Deterministic Control of SDEs with Stochastic Drift and Multiplicative Noise: A Variational Approach

(Article begins on next page)

14 October 2024

DETERMINISTIC CONTROL OF SDES WITH STOCHASTIC DRIFT AND MULTIPLICATIVE NOISE: A VARIATIONAL APPROACH

GIACOMO ASCIONE* AND GIUSEPPE D'ONOFRIO†

Abstract. We consider a linear stochastic differential equation with stochastic drift and multiplicative noise. We study the problem of approximating its solution with the process that solves the equation where the possibly stochastic drift is replaced by a deterministic function. To do this, we use a combination of deterministic Pontryagin's maximum principle approach and direct methods of calculus of variations. We find necessary and sufficient conditions for a function $u \in L^1(0, T)$ to be a minimizer of a certain cost functional. To overcome the problem of the existence of such minimizer, we also consider suitable families of penalized coercive cost functionals. Finally, we consider the important example of the quadratic cost functional, showing that the expected value of the drift component is not always the best choice in the mean squared error approximation.

1. Introduction

Optimal control of dynamical systems consists in the optimization, via a suitable control, of certain measures of performance of the system. Precisely, assuming that the state of the system is described by a differential equation, we want to *modifiy* the equation with a function (called control) belonging to a suitable class in order to minimize a certain functional depending on both the controlled state of the system and the control itself. In the context of stochastic calculus this problem extends naturally to the case in which the system is described through a controlled stochastic differential equation (SDE). Historically, the latter is addressed by two main theoretical approaches that have been developed starting from Bellman's and Pontryagin's optimality principles (see for instance the comprehensive survey by Pham [36]). The first one is called the dynamic programming principle, based on Bellman's optimality principle [8]: it consists in defining a dynamic value function by using the cost functional and then trying to describe it via partial differential equations (PDEs). This method relies on a class of nonlinear PDEs called Hamilton-Jacobi-Bellmann equations [25, 27]. Let us emphasize that one can also adapt the latter to more *complex* situation (e.g. [7]). The second approach, instead, is based on a stochastic generalization of Pontryagin's maximum principle [35]. While the deterministic version can be expressed, in some suitable cases, via a forward-backward differential system, the stochastic one led to the definition of backward stochastic differential equations (BSDEs) [34]. Let us also stress that the stochastic maximum principle usually works with second variations (while the deterministic one only with first) due to the presence of the white noise. This branch of control theory considerably developed over the last years [1, 11, 14, 20, 21, 29, 33, 41]. Here we want to address an approximation problem concerning a linear SDE. Indeed, the tools coming from optimal control theory have been already used to approach some approximation problems. This is done, for instance, in [23] where a stochastic control problem is approximated by a sequence of deterministic control problems, obtaining a Wong-Zakai like ([45]) convergence result. Actually we are interested in approximating an SDE admitting a stochastic drift with another one in which such drift is replaced by a deterministic one.

More precisely, in this paper we consider the following type of linear SDEs

$$(1.1) \qquad \begin{cases} dX(t) = [a(t)X(t) + z(t)]dt + X(t)dW(t), & t \in [0, T] \\ X(0) = X_0, \end{cases}$$

with multiplicative noise and where $z(t)$, appearing in the drift term, is a suitable stochastic process. This kind of equations arises in many applications ranging from finance [32] to neuronal modeling [18, 22] or quickest detection [24]. Moreover, if $z(t)$ is itself the solution of an SDE, Eq.(1.1) plays a role in many systems of equations used in epidemiology, climate models, game theory and others [3, 10, 19].

We are interested in finding the *best* approximation for a solution of Eq.(1.1) obtained by substituting the possibly stochastic drift with a deterministic function. A measure of goodness of the approximation is expressed via the cost functional

$$(1.2) \qquad J : u \in L^1(0,T) \to \mathbb{E}\left[\int_0^T F(t, |X(t) - X_u(t)|)dt\right],$$

where F is a suitably regular function depending on the distance between $X(t)$ and $X_u(t)$, that is the solution of Eq.(1.1) where we replace z by u. Let us underline that the Lagrangian function F does not depend directly on u. Usually, this could lead to a trivial solution of a control problem. Triviality is avoided since we are constraining u to be deterministic. Our aim is to find, if it exists, a function $\overline{u} \in L^1(0,T)$ that minimizes J. In the literature, to the best of our knowledge, few contributions on purely deterministic controls of stochastic equations are available [4, 40].

In [4] we considered the problem of approximating the solution of an SDE with stochastic drift and additive noise through an Ornstein–Uhlenbeck type process, by using direct methods of calculus of variations. Conditions for existence and uniqueness of the approximation and bounds on the goodness of the corresponding approximations are given for some examples. However, in that work, the presence of just additive noise allowed us to reformulate the problem on the class of absolutely continuous functions and led to a purely deterministic treatment. The multiplicative noise, on the other hand, requires a different approach.

Here we find necessary and sufficient conditions for a function u in $L^1(0,T)$ to be a minimizer of J, while we are not able to prove the existence of such a solution in a general setting. To overcome this problem we consider suitable families of penalized cost functionals and we prove that they always admit minimizers. With this property in mind we are able to exploit a sufficient (and necessary) condition for the existence of a solution of the original problem. If the latter condition is not clearly satisfied, then, in any case, the original cost functional evaluated in the solutions of the penalized problems converges towards its infimum as the penalization constant goes to zero. On the other hand, if the condition is satisfied, we can guarantee only weak L^1 convergence of the penalized solution towards the actual solution, but under further regularity assumptions we still have convergence in distribution of the corresponding approximated processes. In the overall, the method we present here can be considered as a combination of deterministic Pontryagin's maximum principle approach and direct methods of calculus of variations.

The paper is structured as follows: in Section 2 we first show some basic properties of the solution of Eq.(1.1) and then we introduce the approximation problem. Section 3 is devoted to obtaining the Euler-Lagrange equation of the functional; i.e we give necessary conditions for a function u to be a minimizer of J. In Section 4 we prove that, under suitable convexity assumptions, the aforementioned Euler-Lagrange equation is also a sufficient condition. In Section 5 we study the penalized problems and we address the problem of existence of a solution and convergence of the penalized solutions to the actual one. Finally, in Section 6, we consider the important example of the quadratic cost functional. While on one hand we are able to show that if z is independent of W a solution exists and it is trivially the expected value of z, on the other hand we also provide an example in which it is **not** a minimizer for the quadratic cost functional. This result can be reformulated saying that, in the multiplicative noise case, the expected value of z is not always the best choice in the mean squared error approximation. Due to the non-trivial nature of the Euler-Lagrange equation, all the examples provided in the section have been obtained by using numerical methods for solution of integral equations via MATLAB R2021a [28].

2. The linear equation with multiplicative noise and the approximation problem

2.1. **The linear equation.** Let us consider a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{F}_t)$ and a $\mathcal{F}_t$-Brownian motion $\{W(t), t \geq 0\}$. Fix $T > 0$ and consider $\{z(t), t \geq 0\}$ a $\mathcal{F}_t$-adapted process such that

(H1) There exists $p \geq 2$ such that for any fixed $t \in [0,T]$, $z(t) \in L^p(\Omega, \mathbb{P})$ and

$$\int_0^T \mathbb{E}[|z(t)|^p]^{\frac{2}{p}}dt < +\infty.$$

Let us stress out that last condition implies, by Hölder inequality, that

$$\int_0^T \mathbb{E}[|z(t)|^2]dt \leq \int_0^T \mathbb{E}[|z(t)|^p]^{\frac{2}{p}}dt < +\infty.$$

On the other hand, let us observe that, denoting $U = \{t \in [0,T] : \mathbb{E}[|z(t)|^p] \geq 1\}$ and $U^c = [0,T] \setminus U$,

$$\int_0^T \mathbb{E}[|z(t)|^p]^{\frac{1}{p}} dt = \int_U \mathbb{E}[|z(t)|^p]^{\frac{1}{p}} dt + \int_{U^c} \mathbb{E}[|z(t)|^p]^{\frac{1}{p}} dt$$

$$\leq \int_U \mathbb{E}[|z(t)|^p]^{\frac{2}{p}} dt + |U^c| \leq \int_0^T \mathbb{E}[|z(t)|^p]^{\frac{2}{p}} dt + T < +\infty.$$

Finally, let us observe that

$$\mathbb{E}\left[\int_0^T |z(t)|dt\right] = \int_0^T \mathbb{E}[|z(t)|]dt \leq \int_0^T \mathbb{E}[|z(t)|^p]^{\frac{1}{p}} dt < +\infty,$$

hence $\int_0^T |z(t)|dt$ is $\mathbb{P}$-almost surely finite and $z \in L^1(0,T)$ $\mathbb{P}$-almost surely. Let us denote by $\mathcal{L}_p^2(\Omega, \mathbb{P}; [0,T])$ the space of $\mathcal{F}_t$-adapted processes $\{z(t),\ t \geq 0\}$ satisfying $(H1)$. The notation is justified by the fact that $(H1)$ can be also written as

$$\left\| \|z(\cdot)\|_{L^p(\Omega,\mathbb{P})} \right\|_{L^2(0,T)} < +\infty.$$

Let us also consider a function $a : [0,T] \to \mathbb{R}$ in $L^\infty(0,T)$. We focus on the linear SDE

(2.1)
$$\begin{cases} dX(t) = [a(t)X(t) + z(t)]dt + X(t)dW(t), & t \in [0,T] \\ X(0) = X_0, \end{cases}$$

where $X_0 \in L^2(\Omega, \mathbb{P})$. In particular the following result holds.

Proposition 2.1. *Let $\mathcal{L}_1^2([0,T];\Omega,\mathbb{P})$ be the space of $\mathcal{F}_t$-adapted processes $\{z(t),\ t \geq 0\}$ such that $z(\cdot) \in L^1(0,T)$ $\mathbb{P}$-almost surely and*

$$\mathbb{E}\left[\left(\int_0^T |z(t)|dt\right)^2\right] < +\infty.$$

Then the map $\mathcal{S}_{X_0} : \mathcal{L}_1^2([0,T];\Omega,\mathbb{P}) \mapsto \mathcal{L}_2^2(\Omega,\mathbb{P};[0,T])$, such that for any $z \in \mathcal{L}_1^2([0,T];\Omega,\mathbb{P})$ the process $\mathcal{S}_{X_0} z$ is solution of (2.1), is well-defined and it holds

(2.2)
$$\mathcal{S}_{X_0} z(t) = G(t)e^{A(t)}\left(X_0 + \int_0^t \frac{e^{-A(s)}}{G(s)} z(s)ds\right), \quad \forall z \in \mathcal{L}_1^2([0,T];\Omega,\mathbb{P}),$$

where $A(t) = \int_0^t a(s)ds$ and $G(t)$ is the geometric Brownian motion associated to $W(t)$, i.e.

(2.3)
$$G(t) = e^{W(t)-\frac{1}{2}t}$$

Proof. By a simple adaptation of the proof of [31, Theorem 5.2.1], the SDE (2.1) admits a unique strong solution in $\mathcal{L}_2^2(\Omega,\mathbb{P};[0,T])$ whenever $z \in \mathcal{L}_1^2([0,T];\Omega,\mathbb{P})$.

Let us prove Equation (2.2). To do this, let us consider the linear SDE

$$\begin{cases} dY(t) = (1 - a(t))Y(t)dt - Y(t)dW(t), & t \in [0,T] \\ Y(0) = 1 \end{cases}$$

and define $Z(t) = \log(Y(t))$. By Itô's formula we have

$$\begin{cases} dZ(t) = \left(\frac{1}{2} - a(t)\right)dt - dW(t), & t \in [0,T] \\ Z(0) = 0 \end{cases}$$

and then, integrating

$$Z(t) = \frac{1}{2}t - W(t) - \int_0^t a(s)ds, \ t \in [0,T].$$

Recalling the definition of $Z(t)$, we have

(2.4)
$$Y(t) = \frac{e^{-A(t)}}{G(t)}, \ t \in [0,T].$$

Let $X(t) = \mathcal{S}_{X_0} z(t)$. By Itô's formula we have

$$d(X(t)Y(t)) = Y(t)dX(t) + X(t)dY(t) - X(t)Y(t)dt = z(t)Y(t)dt.$$

Integrating the previous relation we have

$$X(t)Y(t) = X_0 + \int_0^t Y(s)z(s)ds.$$

Equation (2.4) concludes the proof.   □

Remark 2.2. Since Equation (2.1) is linear, one could explicitly write the solution in terms of $z(t)$ apparently just supposing that $z \in L^1(0,T)$ $\mathbb{P}$-almost surely. However, this does not guarantee that $X(t)$ is regular enough to admit an Itô integral, which is instead needed to express the equation itself.

Let us also observe that $\mathcal{L}_p^2(\Omega, \mathbb{P}; [0,T]) \cup L^1(0,T) \subset \mathcal{L}_1^2([0,T]; \Omega, \mathbb{P})$, where with $L^1(0,T)$ we denote the space of absolutely integrable deterministic functions, considered as degenerate stochastic processes.

As a direct consequence of the previous result we obtain the following.

Corollary 2.3. *The solution map $\mathcal{S}_{X_0}$ is affine, i.e. for any $n \in \mathbb{N}$, $(a_1, \ldots, a_n) \in \mathbb{R}^n$ such that $\sum_{i=1}^n a_i = 1$ and $z_1, \ldots, z_n \in \mathcal{L}_1^2([0,T]; \Omega, \mathbb{P})$ it holds*

$$\mathcal{S}_{X_0}\left(\sum_{i=1}^n a_i z_i\right) = \sum_{i=1}^n a_i \mathcal{S}_{X_0} z_i.$$

Moreover, for any $X_1, X_2 \in L^2(\Omega, \mathbb{P})$ and $z_1, z_2 \in \mathcal{L}_1^2([0,T]; \Omega, \mathbb{P})$ it holds

$$\mathcal{S}_{X_1} z_1 - \mathcal{S}_{X_2} z_2 = \mathcal{S}_{X_1 - X_2}(z_1 - z_2).$$

Finally, $\mathcal{S}_0$ is linear.

Proof. Let us just observe that

$$\mathcal{S}_{X_0}\left(\sum_{i=1}^n a_i z_i\right)(t) = G(t)e^{A(t)}\left(X_0 + \int_0^t \frac{e^{-A(s)}}{G(s)}\left(\sum_{i=1}^n a_i z_i(s)\right)ds\right)$$

$$= G(t)e^{A(t)}\left(\sum_{i=1}^n a_i X_0 + \sum_{i=1}^n a_i \int_0^t \frac{e^{-A(s)}}{G(s)} z_i(s)ds\right)$$

$$= \sum_{i=1}^n a_i G(t)e^{A(t)}\left(X_0 + \int_0^t \frac{e^{-A(s)}}{G(s)} z_i(s)ds\right) = \sum_{i=1}^n a_i \mathcal{S}_{X_0} z_i(t).$$

The second and third statements can be proved in an analogous way.   □

Next, we want to underline some properties of the moments of $\mathcal{S}_{X_0} z(t)$ when z belongs to a certain Banach space. To do this, let us introduce the Banach space $\mathcal{L}_p^1(\Omega, \mathbb{P}; [0,T])$ of the $\mathcal{F}_t$-adapted processes $\{z(t), \ t \geq 0\}$ such that

$$\int_0^T \mathbb{E}[|z(t)|^p]^{\frac{1}{p}} dt < +\infty.$$

Clearly, we have $\mathcal{L}_p^2(\Omega, \mathbb{P}; [0,T]) \cup L^1(0,T) \subset \mathcal{L}_p^1(\Omega, \mathbb{P}; [0,T]) \cap \mathcal{L}_1^2([0,T]; \Omega, \mathbb{P})$. On the other hand, let us also recall the following moment estimate for linear SDEs (see [46, Chapter 3, Lemma 4.2]).

Lemma 2.4. *Consider $\kappa \geq 1$ and let $Y(t)$ be a strong solution of*

$$\begin{cases} dY(t) = [a_1(t)Y(t) + a_2(t)]dt + [b_1(t)Y(t) + b_2(t)]dW(t), & t \in [0,T] \\ Y(0) = Y_0 \end{cases}$$

where $Y_0 \in L^{2\kappa}(\Omega, \mathbb{P})$, $a_1, b_1 : [0,T] \to \mathbb{R}$ are functions in $L^\infty(0,T)$ with $M \geq \max\{\|a_1\|_{L^\infty(0,T)}, \|b_1\|_{L^\infty(0,T)}\}$ and

$$\int_0^T \mathbb{E}[|a_2(t)|^{2\kappa}]^{\frac{1}{2\kappa}} dt + \int_0^T \mathbb{E}[|b_2(t)|^{2\kappa}]^{\frac{1}{\kappa}} dt < +\infty.$$

Then there exists a constant $K(\kappa, M, T) > 0$ such that

$$\sup_{t \in [0,T]} \mathbb{E}[|Y(t)|^{2\kappa}] \leq K(\kappa, M, T)\left(\mathbb{E}[|Y_0|^{2\kappa}] + \left(\int_0^T \mathbb{E}[|a_2(t)|^{2\kappa}]^{\frac{1}{2\kappa}} dt\right)^{2\kappa} + \left(\int_0^T \mathbb{E}[|b_2(t)|^{2\kappa}]^{\frac{1}{\kappa}} dt\right)^\kappa\right).$$

Moreover, for fixed $\kappa \geq 1$ and $M > 0$, the function $T > 0 \mapsto K(\kappa, M, T)$ is increasing.

Remark 2.5. Actually, the last statement of the Lemma is a direct consequence of the constructive proof presented in [46, Chapter 3, Lemma 4.2].

By using the previous Lemma, we have the following result.

Lemma 2.6. *Let* $z \in \mathcal{L}_p^1(\Omega, \mathbb{P}; [0,T]) \cap \mathcal{L}_1^2([0,T]; \Omega, \mathbb{P})$ *and* $X_0 \in L^p(\Omega, \mathbb{P})$ *for some* $p \geq 2$*. Then it holds*

$$\sup_{t \in [0,T]} \mathbb{E}[|\, \mathcal{S}_{X_0}\, z(t)|^p] < +\infty.$$

Moreover, if $X_0 = 0$ *almost surely and* $u \in L^1(0,T)$*, then*

$$\sup_{t \in [0,T]} \mathbb{E}[|\, \mathcal{S}_0\, u(t)|^p] \leq K\left(\frac{p}{2}, M, T\right) \|u\|_{L^1(0,T)}^p,$$

where $M = \|a\|_{L^\infty(0,T)}$ *and* K *is defined in Lemma 2.4.*

Proof. Being $p \geq 2$, we can consider $\kappa = \frac{p}{2} \geq 1$. By using Lemma 2.4 we have

$$\sup_{t \in [0,T]} \mathbb{E}[|\, \mathcal{S}_{X_0}\, z(t)|^p] \leq K\left(\frac{p}{2}, M, T\right)\left(\mathbb{E}[|X_0|^p] + \left(\int_0^T \mathbb{E}[|z(t)|^p]^{\frac{1}{p}}\, dt\right)^p\right) < +\infty.$$

The second part of the statement easily follows by the fact that $\mathbb{E}[|X_0|^p] = 0$ and $\mathbb{E}[|u(t)|^p]^{\frac{1}{p}} = |u(t)|$. $\square$

Remark 2.7. The arguments in the paper can be carried on without the hypothesis (*H1*), but just considering $z \in \mathcal{L}_p^1(\Omega, \mathbb{P}; [0,T]) \cap \mathcal{L}_1^2([0,T]; \Omega, \mathbb{P})$. Here, for the ease of the reader, we will directly consider $z \in \mathcal{L}_p^2(\Omega, \mathbb{P}; [0,T])$.

2.2. **Some properties of the Geometric Brownian Motion.** As we have seen in the previous subsection, the Geometric Brownian Motion $G(t)$ defined in Equation (2.3) will play a major role. Let us first recall that, it being a Doleans-Dade exponential (see [26, Chapter 1]) with $G(0) = 1$, it is a $\mathcal{F}_t$-martingale. On the other hand, we can consider the process

$$G'(t) = \frac{e^{-t}}{G(t)} = e^{-W(t) - \frac{1}{2}t}.$$

It is not difficult to check that $G'(t)$ is still a Geometric Brownian motion (by the fact that $-W(t)$ is still a Brownian motion) and it is given by the Doleans-Dade exponential of $-W(t)$. Thus, in particular, also $G'(t)$ is a $\mathcal{F}_t$-martingale.
Concerning the distribution of $G(t)$, let us call back that it is a log-normal process such that, for fixed $t > 0$, $\log(G(t)) \sim \mathcal{N}\left(-\frac{1}{2}t, t\right)$. By using the formula of the moment generating function of a Gaussian random variable, it is easy to show that, for any $q \geq 0$,

$$(2.5) \qquad \qquad \mathbb{E}[G(t)^q] = e^{\frac{q(q-1)}{2}t}, \qquad t \geq 0.$$

The same relation holds for $G'(t)$. Combining Equation (2.5) and Doob's maximal inequality (see [38, Theorem II.1.7]) we get the following bound on the supremum of G and G'.

Lemma 2.8. *Let* $p_1, p_2 \geq 0$ *and* $T > 0$*. Then there exists a constant* $C(p_1, p_2, T)$ *such that*

$$\mathbb{E}\left[\left(\sup_{t \in [0,T]} G(t)\right)^{p_1}\left(\sup_{t \in [0,T]} G'(t)\right)^{p_2}\right] \leq C(p_1, p_2, T).$$

Proof. By the Cauchy-Schwartz inequality, we have

$$\mathbb{E}\left[\left(\sup_{t \in [0,T]} G(t)\right)^{p_1}\left(\sup_{t \in [0,T]} G'(t)\right)^{p_2}\right] \leq \mathbb{E}\left[\left(\sup_{t \in [0,T]} G(t)\right)^{2p_1}\right]^{\frac{1}{2}} \mathbb{E}\left[\left(\sup_{t \in [0,T]} G'(t)\right)^{2p_2}\right]^{\frac{1}{2}}.$$

Since $2p_1 \geq 0$, we can use Doob's maximal inequality in L^p form to achieve

$$\mathbb{E}\left[\left(\sup_{t \in [0,T]} G(t)\right)^{2p_1}\right] = \mathbb{E}\left[\sup_{t \in [0,T]} G^{2p_1}(t)\right] \leq \left(\frac{2p_1}{2p_1 - 1}\right)^{2p_1} \sup_{t \in [0,T]} \mathbb{E}[G^{2p_1}(t)] = \left(\frac{2p_1}{2p_1 - 1}\right)^{2p_1} e^{p_1(2p_1 - 1)T},$$

where we also used equation (2.5). In the same way we have

$$\mathbb{E}\left[\left(\sup_{t\in[0,T]} G'(t)\right)^{2p_2}\right] \leq \left(\frac{2p_2}{2p_2-1}\right)^{2p_2} e^{p_2(2p_2-1)T}.$$

Setting

$$C(p_1,p_2,T) = \left(\frac{2p_1}{2p_1-1}\right)^{p_1}\left(\frac{2p_2}{2p_2-1}\right)^{p_2} e^{\frac{p_1(2p_1-1)+p_2(2p_2-1)}{2}T}$$

we conclude the proof. $\qquad\square$

From now on, we will use the symbol C to denote a generic positive constant whose value is not important in our arguments. Whenever we need to underline the dependence of C on some parameters $p_1,\ldots,p_n$ we will denote it as $C(p_1,\ldots,p_n)$. The only exception is Theorem 5.15, in which the constants are indexed to keep track of the dependence on the involved parameters.

2.3. **The approximation problem.** We are interested in finding the *best approximation* for a solution of Equation (2.1), obtained by substituting the possibly stochastic drift with a deterministic function. To do this, let us first introduce a cost functional

$$J : u \in L^1(0,T) \to \mathbb{E}\left[\int_0^T F(t,\xi_u(t))dt\right],$$

where F is a suitable function and $\xi_u = \mathcal{S}_0(z-u)$. Let us consider the following assumptions on F:

(H2) It holds $F(t,\xi) \geq 0$ for any $t \in [0,T]$ and $\xi \in \mathbb{R}$;
(H3) $F(t,\xi)$ is twice continuously differentiable in the ξ variable and $\frac{\partial F}{\partial \xi}(t,\xi)$ and $\frac{\partial^2 F}{\partial \xi^2}(t,\xi)$ are continuous functions of both variables;
(H4) There exist $\alpha \in (0,p)$ and a non-negative function $L \in L^1(0,T)$ such that

$$|F(t,\xi)| + \left|\frac{\partial F}{\partial \xi}(t,\xi)\right| + \left|\frac{\partial^2 F}{\partial \xi^2}(t,\xi)\right| \leq L(t)(1+|\xi|^\alpha), \; t \in [0,T], \; \xi \in \mathbb{R}.$$

Our aim is to find, if it exists, a function $\overline{u} \in L^1(0,T)$ such that

$$J[\overline{u}] = \min_{u\in L^1(0,T)} J[u].$$

We can consider the functional $J[u]$ to be a cost functional for an approximation problem. Indeed, we want to find a deterministic function $u(t)$ that we can substitute to the process $z(t)$ in $X(t) = \mathcal{S}_{X_0} z(t)$ to obtain *the best possible approximation* under the cost J. For this reason we expect the cost functional to depend in some sense on the gap between $X(t)$ and the approximating process $X_u(t) = \mathcal{S}_{X_0} u(t)$. By affinity of the solution map, we have that $\xi_u(t) := X(t) - X_u(t) = \mathcal{S}_0(z-u)(t)$. With this idea in mind, the function F can be seen as a running cost.

Hypothesis (H2) is natural as we want to consider $J[u]$ as a cost functional for an approximation problem, while (H3) is just a regularity assumption. Hypothesis (H4) implies some form of controlled growth for both the running cost F and its first and second derivatives with respect to the gap process ξ_u. The growth assumption on F can be justified by means of the following non-triviality result.

Lemma 2.9. *For any $u \in L^1(0,T)$ it holds $J[u] < +\infty$.*

Proof. We have

$$J[u] = \mathbb{E}\left[\int_0^T F(t,\xi_u(t))dt\right] \leq \int_0^T L(t)(1+\mathbb{E}[|\xi_u(t)|^\alpha])dt.$$

Now let us estimate $\mathbb{E}[|\xi_u(t)|^\alpha]$. To do this, let us consider $\widetilde{p} = \frac{p}{\alpha} > 1$ and let us apply Hölder inequality to achieve

$$\mathbb{E}[|\xi_u(t)|^\alpha] \leq \mathbb{E}[|\xi_u(t)|^p]^{\frac{\alpha}{p}}.$$

Let $C = \sup_{t\in[0,T]} \mathbb{E}[|\xi_u(t)|^p]$, that is finite by Lemma 2.6. Then we have

$$\mathbb{E}[|\xi_u(t)|^\alpha] \leq C^{\frac{\alpha}{p}}.$$

Thus we have

$$J[u] \leq (1 + K^{\frac{\alpha}{p}}) \int_0^T L(t)dt < +\infty,$$

being $L \in L^1(0,T)$. □

The previous result and hypothesis $(H2)$ guarantee that

$$\inf_{u \in L^1(0,T)} J[u] \in [0, +\infty)$$

thus it makes sense to search for a minimizer (if it exists) of $J[u]$. Next section will clarify the role of the first and second derivatives in hypothesis $(H4)$.

3. NECESSARY OPTIMALITY CONDITIONS

Now let us focus on necessary optimality conditions, i.e. conditions that a global minimizer $\overline{u} \in L^1(0,T)$ of the functional $J[u]$ has to satisfy. Let us stress out that, in order to discuss necessary optimality conditions, we assume that we already have a minimizer $\overline{u}$. In particular, necessary optimality conditions are needed to find at least a set of candidate minimizers.

Before going into details, let us introduce some notation. Let $f \in L^1(0,T)$. We denote the set of its Lebesgue points as E_f (see [17, Section 1.7]). From now on, since $f \in L^1(0,T)$ is almost everywhere finite, we will always consider a version that is everywhere finite, so that, for each $t \in E_f$, it holds

$$\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \int_{t-\frac{\varepsilon}{2}}^{t+\frac{\varepsilon}{2}} f(\tau)d\tau = f(t)$$

and we set, for each $t \notin E_f$, $f(t) = 0$. Such version of $f(t)$ is called precise representative of f. To obtain necessary optimality conditions we need the following property:

Proposition 3.1. *Let $f \in L^1(0,T)$ and $g : [0,T] \to \mathbb{R}$ be a continuous function. Define $h(t) = g(t)f(t)$ for any $t \in [0,T]$. Then $h \in L^1(0,T)$ and $E_f \subseteq E_h$.*

The previous statement is classical, but, for completeness, we add its proof in Appendix A.
Now we are ready to prove the main result of this section.

Theorem 3.2. *Suppose hypotheses (H1) to (H4) are satisfied. Let $\overline{u} \in L^1(0,T)$ be a global minimum of the functional J over $L^1(0,T)$. Then it holds*

$$(3.1) \qquad \int_{t_0}^T \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_u(t))\frac{G(t)}{G(t_0)}e^{(A(t)-A(t_0))}\right] dt = 0, \ t_0 \in [0,T].$$

Proof. Let us consider $E_{\overline{u}}$ the set of Lebesgue points of $\overline{u}$ in $(0,T)$, E_L the set of Lebesgue points of L in $(0,T)$, $E = E_{\overline{u}} \cup E_L$ and let $t_0 \in E$. Fix any real number $u \in \mathbb{R}$ and $\varepsilon_0 > 0$ small enough to have $\left(t_0 - \frac{\varepsilon_0}{2}, t_0 + \frac{\varepsilon_0}{2}\right) \subset (0,T)$. Now let us define, for any $\varepsilon \in (0, \varepsilon_0)$, $I_\varepsilon := \left(t_0 - \frac{\varepsilon}{2}, t_0 + \frac{\varepsilon}{2}\right)$, the following needle variation

$$u_\varepsilon(t) = \begin{cases} u & t \in I_\varepsilon, \\ \overline{u}(t) & \text{otherwise.} \end{cases}$$

and, denoting $\xi_\varepsilon(t) := \xi_{u_\varepsilon}(t)$, the value of the cost corresponding to a certain choice of ε

$$g(\varepsilon) = J[u_\varepsilon] = \mathbb{E}\left[\int_0^T F(t, \xi_\varepsilon(t))dt\right].$$

We want to show that g is right differentiable in 0.
Let us consider the auxiliary function

$$h : \theta \in (0,1) \mapsto F(t, \theta\xi_{\overline{u}}(t) + (1-\theta)\xi_\varepsilon(t))$$

for fixed $t \in [0,T]$ and $\varepsilon \in (-\varepsilon_0, \varepsilon_0)$. Let us observe that, by hypothesis $(H3)$, h is twice differentiable and

$$h'(\theta) = \frac{\partial F}{\partial \xi}(t, \theta\xi_{\overline{u}}(t) + (1-\theta)\xi_\varepsilon(t))(\xi_{\overline{u}}(t) - \xi_\varepsilon(t))$$

$$h''(\theta) = \frac{\partial^2 F}{\partial \xi^2}(t, \theta\xi_{\overline{u}}(t) + (1-\theta)\xi_\varepsilon(t))(\xi_{\overline{u}}(t) - \xi_\varepsilon(t))^2.$$

By using the Fundamental Theorem of Calculus and then integrating by parts, we have

$$h(0) - h(1) = -\int_0^1 h'(\theta)d\theta = -h'(1) + \int_0^1 \theta h''(\theta)d\theta$$

that is to say

$$F(t, \xi_\varepsilon(t)) - F(t, \xi_{\overline{u}}(t)) = \frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t))(\xi_\varepsilon(t) - \xi_{\overline{u}}(t))$$

$$+ \int_0^1 \theta \frac{\partial^2 F}{\partial \xi^2}(t, \theta \xi_{\overline{u}}(t) + (1-\theta)\xi_\varepsilon(t))(\xi_{\overline{u}}(t) - \xi_\varepsilon(t))^2 d\theta.$$

We can use the previous representation of $F(t, \xi_\varepsilon(t))$ to rewrite the incremental ratio of g in 0. Indeed, denoting $\eta_\varepsilon(t) := \xi_{\overline{u}}(t) - \xi_\varepsilon(t)$, we get

(3.2)
$$\frac{g(\varepsilon) - g(0)}{\varepsilon} = \frac{1}{\varepsilon} \mathbb{E}\left[\int_0^T \left(-\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t))\eta_\varepsilon(t)\right.\right.$$
$$\left.\left.+ \int_0^1 \theta \frac{\partial^2 F}{\partial \xi^2}(t, \theta \xi_{\overline{u}}(t) + (1-\theta)\xi_\varepsilon(t))\eta_\varepsilon^2(t)d\theta\right)dt\right].$$

By the properties of the solution map $\mathcal{S}_0$ given in Corollary 2.3, we have that

$$\eta_\varepsilon(t) = \mathcal{S}_0(u_\varepsilon - \bar{u})(t) = \begin{cases} 0 & t \le t_0 - \frac{\varepsilon}{2} \\ G(t)e^{A(t)}\int_{t_0-\frac{\varepsilon}{2}}^t \frac{e^{-A(s)}}{G(s)}[u - \overline{u}(s)]ds & t \in I_\varepsilon \\ G(t)e^{A(t)}\int_{I_\varepsilon} \frac{e^{-A(s)}}{G(s)}[u - \overline{u}(s)]ds & t \ge t_0 + \frac{\varepsilon}{2} \end{cases}$$

On the other hand, let us observe that

$$\|u_\varepsilon - \bar{u}\|_{L^1(0,T)} = \int_{I_\varepsilon} |u - \bar{u}(\tau)|d\tau$$

and then, by the second part of Lemma 2.6 we have, for any exponent $\kappa \ge 2$,

(3.3)
$$\sup_{t \in [0,T]} \mathbb{E}[|\eta_\varepsilon(t)|^\kappa] \le C \left(\int_{I_\varepsilon} |u - \bar{u}(\tau)|d\tau\right)^\kappa.$$

Going back to Equation (3.2), let us split the integral as

(3.4)
$$\frac{g(\varepsilon) - g(0)}{\varepsilon} = -\frac{1}{\varepsilon} \mathbb{E}\left[\int_{I_\varepsilon} \frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t))\eta_\varepsilon(t)dt\right]$$
$$-\frac{1}{\varepsilon} \mathbb{E}\left[\int_{t_0+\frac{\varepsilon}{2}}^T \frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t))\eta_\varepsilon(t)dt\right]$$
$$+\frac{1}{\varepsilon} \mathbb{E}\left[\int_0^T \int_0^1 \theta \frac{\partial^2 F}{\partial \xi^2}(t, \theta \xi_{\overline{u}}(t) + (1-\theta)\xi_\varepsilon(t))\eta_\varepsilon^2(t)d\theta dt\right]$$
$$=: I_1(\varepsilon) + I_2(\varepsilon) + I_3(\varepsilon),$$

where we used the fact that $\eta_\varepsilon(t) = 0$ as $t \le t_0 - \frac{\varepsilon}{2}$. Now we want to take the limit as $\varepsilon \to 0$. First of all, let us show that $\lim_{\varepsilon \to 0} I_1(\varepsilon) = 0$. To do this, let us observe that

$$|I_1(\varepsilon)| \le \frac{1}{\varepsilon} \mathbb{E}\left[\int_{I_\varepsilon} \left|\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t))\eta_\varepsilon(t)\right| dt\right]$$
$$= \frac{1}{\varepsilon} \int_{I_\varepsilon} \mathbb{E}\left[\left|\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t))\eta_\varepsilon(t)\right|\right] dt.$$

Bringing back the exponent $\alpha \in (0, p)$ in hypothesis $(H4)$, let us consider any $\widetilde{p} \in \left(1, \min\left\{2, \frac{p}{\alpha}\right\}\right)$. Let $\widetilde{q}$ be its conjugate exponent, i.e. such that $\frac{1}{\widetilde{p}} + \frac{1}{\widetilde{q}} = 1$. Being $\widetilde{p} < 2$ we have $\widetilde{q} > 2$. By Hölder's inequality it holds

(3.5)
$$|I_1(\varepsilon)| \le \frac{1}{\varepsilon} \int_{I_\varepsilon} \left(\mathbb{E}\left[\left|\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t))\right|^{\widetilde{p}}\right]\right)^{\frac{1}{\widetilde{p}}} \left(\mathbb{E}\left[|\eta_\varepsilon(t)|^{\widetilde{q}}\right]\right)^{\frac{1}{\widetilde{q}}} dt.$$

Using the growth hypothesis $(H4)$ we have

$$\mathbb{E}\left[\left|\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t))\right|^{\widetilde{p}}\right] \leq L^{\widetilde{p}}(t)\,\mathbb{E}\left[(1 + |\xi_{\overline{u}}(t)|^{\alpha})^{\widetilde{p}}\right]$$

$$\leq L^{\widetilde{p}}(t)2^{\widetilde{p}-1}(1 + \mathbb{E}[|\xi_{\overline{u}}(t)|^{\alpha\widetilde{p}}])$$

$$\leq L^{\widetilde{p}}(t)2^{\widetilde{p}-1}(1 + (\mathbb{E}[|\xi_{\overline{u}}(t)|^{p}])^{\frac{\alpha\widetilde{p}}{p}}),$$

where we used the convexity of the function $x \mapsto x^{\widetilde{p}}$ and applied Hölder's inequality a second time with the exponent $\frac{p}{\alpha\widetilde{p}} > 1$. Taking the supremum as $t \in [0, T]$ on the right-hand side and using Lemma 2.6 we finally achieve

$$(3.6) \qquad \mathbb{E}\left[\left|\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t))\right|^{\widetilde{p}}\right] \leq CL^{\widetilde{p}}(t)$$

where C is a positive constant. On the other hand, by Equation (3.3), we get

$$(3.7) \qquad \left(\mathbb{E}\left[|\eta_{\varepsilon}(t)|^{\widetilde{q}}\right]\right)^{\frac{1}{\widetilde{q}}} \leq C\left(\int_{I_{\varepsilon}} |u - \overline{u}(\tau)|d\tau\right).$$

Combining equations (3.5), (3.6) and (3.7) we get

$$|I_1(\varepsilon)| \leq C\varepsilon\left(\frac{1}{\varepsilon}\int_{I_{\varepsilon}} L(t)dt\right)\left(\frac{1}{\varepsilon}\int_{I_{\varepsilon}} |u - \overline{u}(t)|dt\right).$$

It is not difficult to see that if $t_0 \in E_{\overline{u}}$, then it is also a Lebesgue point for $|u - \overline{u}(t)|$. Thus, being $t_0 \in E$, we conclude that $\lim_{\varepsilon \to 0} I_1(\varepsilon) = 0$.

Now let us show that $\lim_{\varepsilon \to 0} I_3(\varepsilon) = 0$. Arguing as before we have

$$(3.8) \qquad \begin{aligned} |I_3(\varepsilon)| &\leq \frac{1}{\varepsilon}\int_0^T \int_0^1 \theta\,\mathbb{E}\left[\left|\frac{\partial^2 F}{\partial \xi^2}(t, \theta\xi_{\overline{u}}(t) + (1-\theta)\xi_{\varepsilon}(t))\eta_{\varepsilon}^2(t)\right|\right]d\theta dt \\ &\leq \frac{1}{\varepsilon}\int_0^T \int_0^1 \theta\left(\mathbb{E}\left[\left|\frac{\partial^2 F}{\partial \xi^2}(t, \theta\xi_{\overline{u}}(t) + (1-\theta)\xi_{\varepsilon}(t))\right|^{\widetilde{p}}\right]\right)^{\frac{1}{\widetilde{p}}}\left(\mathbb{E}\left[|\eta_{\varepsilon}(t)|^{2\widetilde{q}}\right]\right)^{\frac{1}{\widetilde{q}}}d\theta dt, \end{aligned}$$

where we used Hölder's inequality with the exponent $\widetilde{p}$. Again, using Hypothesis $(H4)$,

$$\mathbb{E}\left[\left|\frac{\partial^2 F}{\partial \xi^2}(t, \theta\xi_{\overline{u}}(t) + (1-\theta)\xi_{\varepsilon}(t))\right|^{\widetilde{p}}\right] \leq L^{\widetilde{p}}(t)\,\mathbb{E}[(1 + |\theta\xi_{\overline{u}}(t) + (1-\theta)\xi_{\varepsilon}(t)|^{\alpha})^{\widetilde{p}}]$$

$$\leq L^{\widetilde{p}}(t)2^{\widetilde{p}-1}(1 + \mathbb{E}[|\theta\xi_{\overline{u}}(t) + (1-\theta)\xi_{\varepsilon}(t)|^{\alpha\widetilde{p}}])$$

$$\leq L^{\widetilde{p}}(t)2^{\widetilde{p}-1}(1 + \mathbb{E}[|\theta\xi_{\overline{u}}(t) + (1-\theta)\xi_{\varepsilon}(t)|^{p}]^{\frac{\alpha\widetilde{p}}{p}}),$$

where we also used the convexity of the function $x \mapsto |x|^{\widetilde{p}}$ and Hölder's inequality with the exponent $\frac{p}{\alpha\widetilde{p}} > 1$. Moreover, by the properties of the solution map $\mathcal{S}_0$ as in Corollary 2.3,

$$\theta\xi_{\overline{u}}(t) + (1-\theta)\xi_{\varepsilon}(t) = \mathcal{S}_0(\theta\overline{u} + (1-\theta)u_{\varepsilon})(t)$$

and then, by Lemma 2.6, we conclude

$$(3.9) \qquad \begin{aligned} \mathbb{E}\left[\left|\frac{\partial^2 F}{\partial \xi^2}(t, \theta\xi_{\overline{u}}(t) + (1-\theta)\xi_{\varepsilon}(t))\right|^{\widetilde{p}}\right] &\leq L^{\widetilde{p}}(t)\,\mathbb{E}[(1 + |\theta\xi_{\overline{u}}(t) + (1-\theta)\xi_{\varepsilon}(t)|^{\alpha})^{\widetilde{p}}] \\ &\leq L^{\widetilde{p}}(t)2^{\widetilde{p}-1}(1 + \mathbb{E}[|\theta\xi_{\overline{u}}(t) + (1-\theta)\xi_{\varepsilon}(t)|^{\alpha\widetilde{p}}]) \\ &\leq CL^{\widetilde{p}}(t). \end{aligned}$$

On the other hand, by Equation (3.3) we know that

$$(3.10) \qquad \left(\mathbb{E}\left[|\eta_{\varepsilon}(t)|^{2\widetilde{q}}\right]\right)^{\frac{1}{\widetilde{q}}} \leq K\left(\int_{I_{\varepsilon}} |u - \overline{u}(\tau)|d\tau\right)^2.$$

Combining Equations (3.8), (3.10) and (3.9) we obtain

$$|I_3(\varepsilon)| \leq \frac{C}{\varepsilon} \left(\int_{I_\varepsilon} |u - \bar{u}(\tau)| d\tau \right)^2 \int_0^T \int_0^1 \theta L(t) dt \leq C\varepsilon \|L\|_{L^1(0,T)} \left(\frac{1}{\varepsilon} \int_{I_\varepsilon} |u - \bar{u}(\tau)| d\tau \right)^2.$$

Taking the limit as $\varepsilon \to 0$ we conclude that $\lim_{\varepsilon \to 0} I_3(\varepsilon) = 0$.

Finally, we need to evaluate $\lim_{\varepsilon \to 0} I_2(\varepsilon)$. To do this, let us first show that we can use Fubini's theorem to exchange the order of expectation and Lebesgue integral. Indeed we have

$$\int_{t_0}^T \mathbb{E}\left[\left| \frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t))\eta_\varepsilon(t) \right| \right] \mathbf{1}_{[t_0+\frac{\varepsilon}{2}, T]}(t) dt \leq \int_{t_0}^T \left(\mathbb{E}\left[\left| \frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) \right|^{\widetilde{p}} \right] \right)^{\frac{1}{\widetilde{p}}} \mathbb{E}\left[|\eta_\varepsilon(t)|^{\widetilde{q}} \right]^{\frac{1}{\widetilde{q}}} dt$$

$$\leq C \left(\int_{I_\varepsilon} |u - \bar{u}(\tau)| d\tau \right) \|L\|_{L^1(0,T)}$$

$$\leq C \|u - \bar{u}\|_{L^1(0,T)} \|L\|_{L^1(0,T)},$$

where, for any $B \subseteq [0, T]$, $\mathbf{1}_B$ is the indicator function of the set B and we used again Equations (3.6) and (3.7). Hence, by Fubini's theorem, we have

$$
\begin{aligned}
(3.11) \quad I_2(\varepsilon) &= -\frac{1}{\varepsilon} \int_{t_0}^T \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t))\eta_\varepsilon(t) \right] \mathbf{1}_{[t_0+\frac{\varepsilon}{2}, T]}(t) dt \\
&= -\int_{t_0}^T \mathbb{E}\left[\mathbf{1}_{[t_0+\frac{\varepsilon}{2}, T]}(t) \frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) G(t) e^{A(t)} \frac{1}{\varepsilon} \int_{I_\varepsilon} \frac{e^{-A(s)}}{G(s)} [u - \overline{u}(s)] ds \right] dt,
\end{aligned}
$$

where we explicitly wrote $\eta_\varepsilon(t)$. Now let us show that we can take the limit inside both the integral and the expectation sign. To do this, we want to use dominated convergence theorem. Let us observe that

$$\left| \mathbf{1}_{[t_0+\frac{\varepsilon}{2}, T]}(t) \frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) G(t) e^{A(t)} \frac{1}{\varepsilon} \int_{I_\varepsilon} \frac{e^{-A(s)}}{G(s)} [u - \overline{u}(s)] ds \right|$$

$$\leq e^T \left(\sup_{t \in [0,T]} e^{2A(t)} \right) \left(\sup_{t \in [0,T]} G(t) \right) \left(\sup_{t \in [0,T]} G'(t) \right) \left(\frac{1}{\varepsilon} \int_{I_\varepsilon} |u - \bar{u}(s)| ds \right) \left| \frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) \right|$$

$$= C \left(\sup_{t \in [0,T]} G(t) \right) \left(\sup_{t \in [0,T]} G'(t) \right) \left(\frac{1}{\varepsilon} \int_{I_\varepsilon} |u - \bar{u}(s)| ds \right) \left| \frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) \right|.$$

In particular we have

$$\lim_{\varepsilon \to 0} \left(\frac{1}{\varepsilon} \int_{I_\varepsilon} |u - \bar{u}(s)| ds \right) = |u - \bar{u}(t_0)|,$$

hence we can suppose ε is small enough to have

$$\left(\frac{1}{\varepsilon} \int_{I_\varepsilon} |u - \bar{u}(s)| ds \right) \leq 2|u - \bar{u}(t_0)|.$$

This implies

$$
\begin{aligned}
(3.12) \quad &\left| \mathbf{1}_{[t_0+\frac{\varepsilon}{2}, T]}(t) \frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) G(t) e^{A(t)} \frac{1}{\varepsilon} \int_{I_\varepsilon} \frac{e^{-A(s)}}{G(s)} [u - \overline{u}(s)] ds \right| \\
&\leq C \left(\sup_{t \in [0,T]} G(t) \right) \left(\sup_{t \in [0,T]} G'(t) \right) \left| \frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) \right|.
\end{aligned}
$$

Now let us show that the stochastic process on the right-hand side is integrable. Observe that

$$\mathbb{E}\left[\left(\sup_{t \in [0,T]} G(t) \right) \left(\sup_{t \in [0,T]} G'(t) \right) \left| \frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) \right| \right]$$

$$\leq \mathbb{E}\left[\left(\sup_{t \in [0,T]} G(t) \right)^{\widetilde{q}} \left(\sup_{t \in [0,T]} G'(t) \right)^{\widetilde{q}} \right]^{\frac{1}{\widetilde{q}}} \mathbb{E}\left[\left| \frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) \right|^{\widetilde{p}} \right]^{\frac{1}{\widetilde{p}}} \leq CL(t),$$

where we used Equation (3.6) and Lemma 2.8. Integrating on $[t_0, T]$ we conclude that

$$\int_{t_0}^{T} \mathbb{E}\left[\left(\sup_{t \in [0,T]} G(t)\right)\left(\sup_{t \in [0,T]} G'(t)\right)\left|\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t))\right|\right] \leq C \left\|L\right\|_{L^1(0,T)}.$$

We only need to show that the integrand in Equation (3.11) converges almost everywhere. Recalling that $t \mapsto G(t)$ is almost surely continuous, fix $\omega \in \Omega$ such that $G(\cdot, \omega)$ is a continuous function, then $t_0 \in E$ is a Lebesgue point for $\frac{e^{-A(t)}}{G(t,\omega)}[u - \overline{u}(t)]$ by Proposition 3.1. Hence we have

$$\lim_{\varepsilon \to 0} \mathbf{1}_{[t_0 + \frac{\varepsilon}{2}, T]}(t) \frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) G(t) e^{A(t)} \frac{1}{\varepsilon} \int_{I_\varepsilon} \frac{e^{-A(s)}}{G(s)}[u - \overline{u}(s)] ds$$

$$= \frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) G(t) e^{A(t)} \frac{e^{-A(t_0)}}{G(t_0)}[u - \overline{u}(t_0)] \text{ a.s.}$$

Thus, by Dominated Convergence Theorem, we get

$$\lim_{\varepsilon \to 0} I_2(\varepsilon) = -(u - \overline{u}(t_0)) \int_{t_0}^{T} \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) e^{A(t) - A(t_0)} \frac{G(t)}{G(t_0)}\right] dt.$$

In conclusion, from Equation (3.4) we have

$$(3.13) \qquad \lim_{\varepsilon \to 0} \frac{g(\varepsilon) - g(0)}{\varepsilon} = -(u - \overline{u}(t_0)) \int_{t_0}^{T} \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) e^{A(t) - A(t_0)} \frac{G(t)}{G(t_0)}\right] dt.$$

However, by definition of $\overline{u}$, we know that 0 is a minimum point for g, and then

$$\lim_{\varepsilon \to 0} \frac{g(\varepsilon) - g(0)}{\varepsilon} \geq 0,$$

that implies

$$u \int_{t_0}^{T} \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) e^{A(t) - A(t_0)} \frac{G(t)}{G(t_0)}\right] dt \leq \overline{u}(t_0) \int_{t_0}^{T} \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) e^{A(t) - A(t_0)} \frac{G(t)}{G(t_0)}\right] dt.$$

Define $H(u) := u \int_{t_0}^{T} \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) e^{A(t) - A(t_0)} \frac{G(t)}{G(t_0)}\right] dt$ and observe that $\overline{u}(t_0)$ is a maximum point for $H(u)$. Hence, by Fermat's theorem, we get $H'(\overline{u}(t_0)) = 0$, that is to say, being $t_0 \in E$ arbitrary,

$$(3.14) \qquad \int_{t_0}^{T} \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) e^{A(t) - A(t_0)} \frac{G(t)}{G(t_0)}\right] dt = 0, \ \forall t_0 \in E.$$

Now we want to extend E to the whole interval $[0, T]$. Let us show that

$$(3.15) \qquad t_0 \in [0, T] \to \int_{t_0}^{T} \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) e^{A(t) - A(t_0)} \frac{G(t)}{G(t_0)}\right] dt$$

is continuous. To do this, consider $t_1 \in [0, T]$ and $t_2 = t_1 + \delta$ for some δ small enough to have $t_2 \in [0, T]$. To fix the ideas, let us suppose $\delta > 0$, since the arguments for $\delta < 0$ are the same. We have

$$\left|\int_{t_1}^{T} \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) e^{A(t) - A(t_1)} \frac{G(t)}{G(t_1)}\right] dt - \int_{t_2}^{T} \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) e^{A(t) - A(t_2)} \frac{G(t)}{G(t_2)}\right] dt\right|$$

$$\leq \int_{t_1}^{t_2} \mathbb{E}\left[\left|\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) e^{A(t) - A(t_1)} \frac{G(t)}{G(t_1)}\right|\right] dt$$

$$+ \int_{t_2}^{T} \mathbb{E}\left[\left|\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) e^{A(t) - A(t_1)} \frac{G(t)}{G(t_1)} - \frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) e^{A(t) - A(t_2)} \frac{G(t)}{G(t_2)}\right|\right] dt$$

$$:= I_4(\delta) + I_5(\delta).$$

Let us consider $I_4(\delta)$. The exact same argument we considered for $I_2(\varepsilon)$ leads to

$$\mathbb{E}\left[\left|\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) e^{A(t) - A(t_1)} \frac{G(t)}{G(t_1)}\right|\right] \leq CL(t)$$

and then

$$I_4(\delta) \leq C \int_{t_1}^{t_2} L(t)dt.$$

By absolute continuity of the Lebesgue integral, we have $\lim_{\delta \to 0} I_4(\delta) = 0$.

Concerning $I_5(\delta)$, we have to use dominated convergence theorem. To do this, let us just observe that, as in Equation (3.12),

$$\left| \frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) e^{A(t)-A(t_1)} \frac{G(t)}{G(t_1)} \right| \leq C \left| \frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}}(t)) \right| \left(\sup_{t \in [0,T]} G(t) \right) \left(\sup_{t \in [0,T]} G'(t) \right)$$

where the right-hand side is independent of t_1 and integrable. The same can be done with t_2 and then, by triangular inequality, we have that the integrand in $I_5(\delta)$ is dominated. By dominated convergence theorem, since $G(t)$ is almost surely continuous, we have $\lim_{\delta \to 0} I_5(\delta) = 0$.

Hence, the function in Equation (3.15) is continuous and, since $|[0,T] \setminus E| = 0$ and thus E is dense in $[0,T]$, we can extend Equation (3.14) to the whole interval $[0,T]$, concluding the proof. $\square$

Remark 3.3. Let us observe that, with the same arguments, we can actually show that J is Gateaux differentiable (see [9]) in any $u \in L^1(0,T)$ with Gateaux derivative given by

$$\partial_u J[v] = -\int_0^T v(s) \int_s^T \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_u(t)) e^{A(t)-A(s)} \frac{G(t)}{G(s)} \right] dt ds, \ v \in L^1(0,T)$$

and then Equation (3.1) can be restated as

$$\partial_{\bar{u}} J[v] = 0, \ \forall v \in L^1(0,T).$$

From this point of view, Equation (3.1) is a consequence of Fermat's theorem applied directly on J. For this reason, we can refer to Equation (3.1) as the Euler-Lagrange equation for the functional J. In the same fashion, we can recognize H as the Hamiltonian function of the cost functional J.

Moreover, let us stress out that, by the absolute continuity of Lebesgue's integral, for any $\delta > 0$ there exists $\varepsilon_0 > 0$ such that for any $\varepsilon < \varepsilon_0$ it holds

$$\|u_\varepsilon - \bar{u}\|_{L^1} = \int_{I_\varepsilon} |u - \bar{u}(\tau)| d\tau < \delta,$$

being $|I_\varepsilon| < \varepsilon_0$. Thus, defining the ball

$$B_\delta(\bar{u}) = \{u \in L^1(0,T): \ \|u - \bar{u}\|_{L^1} < \delta\}$$

we know that, for any fixed $\delta > 0$, there exists $\varepsilon_0 > 0$ such that $u_\varepsilon \in B_\delta(\bar{u})$ for any $\varepsilon < \varepsilon_0$. Last observation leads to the fact that Theorem 3.1 holds also for local minimizers of J, i.e. for functions $\bar{u}$ for which there exists $\delta > 0$ such that

$$J[u] \geq J[\bar{u}], \ \forall u \in B_\delta(\bar{u}).$$

4. Sufficient optimality conditions

In the previous section we obtained a necessary optimality condition given in terms of Equation (3.1). Now we want to investigate whether such condition is also sufficient, i.e. any solution of Equation (3.1) is actually a minimizer of the functional J on $L^1(0,T)$. To do this, we need some additional hypotheses:

(H5) For any fixed $t \in [0,T]$, the function $x \mapsto F(t,x)$ is convex.
(H5+) For any fixed $t \in [0,T]$, the function $x \mapsto F(t,x)$ is strictly convex.

Theorem 4.1. *Suppose Hypotheses (H1) to (H5) are satisfied. Let $\bar{u} \in L^1(0,T)$ be a solution of Equation (3.1). Then $\bar{u}$ is a global minimizer of J.*

Proof. Let us consider $\bar{u} \in L^1(0,T)$ solution of Equation (3.1) and let $u \in L^1(0,T)$ be any other function. Then we have

$$J[u] - J[\bar{u}] = \int_0^T \mathbb{E}[F(t, \xi_u(t)) - F(t, \xi_{\bar{u}}(t))]dt$$

where we already used Fubini's theorem, by means of hypotheses $(H2)$. By hypothesis $(H3)$ and $(H5)$ we have

$$
\begin{aligned}
F(t,\xi_u(t)) - F(t,\xi_{\bar{u}}(t)) &\geq \frac{\partial F}{\partial \xi}(t,\xi_{\bar{u}}(t))(\xi_u(t) - \xi_{\bar{u}}(t)) \\
&= \frac{\partial F}{\partial \xi}(t,\xi_{\bar{u}}(t))\,\mathcal{S}_0(\bar{u} - u) \\
&= \frac{\partial F}{\partial \xi}(t,\xi_{\bar{u}}(t))G(t)e^{A(t)}\int_0^t \frac{e^{-A(s)}}{G(s)}(\bar{u}(s) - u(s))ds \\
&= \int_0^t \frac{\partial F}{\partial \xi}(t,\xi_{\bar{u}}(t))e^{A(t)-A(s)}\frac{G(t)}{G(s)}(\bar{u}(s) - u(s))ds
\end{aligned}
$$

and then

$$
(4.1) \qquad J[u] - J[\bar{u}] \geq \int_0^T \mathbb{E}\left[\int_0^t \frac{\partial F}{\partial \xi}(t,\xi_{\bar{u}}(t))e^{A(t)-A(s)}\frac{G(t)}{G(s)}(\bar{u}(s) - u(s))ds\right]dt.
$$

Now we want to exchange the order of the integrals. To do this, observe that

$$
\begin{aligned}
&\left|\frac{\partial F}{\partial \xi}(t,\xi_{\bar{u}}(t))e^{A(t)-A(s)}\frac{G(t)}{G(s)}(\bar{u}(s) - u(s))\right| \\
&\leq \left|\frac{\partial F}{\partial \xi}(t,\xi_{\bar{u}}(t))\right|\left(\sup_{\tau_1,\tau_2\in[0,T]}e^{A(\tau_1)-A(\tau_2)}\right)e^T\left(\sup_{\tau\in(0,T)}G(\tau)\right)\left(\sup_{\tau\in(0,T)}G'(\tau)\right)|\bar{u}(s) - u(s)| \\
&= C\left|\frac{\partial F}{\partial \xi}(t,\xi_{\bar{u}}(t))\right|\left(\sup_{\tau\in(0,T)}G(\tau)\right)\left(\sup_{\tau\in(0,T)}G'(\tau)\right)|\bar{u}(s) - u(s)|.
\end{aligned}
$$

Let us consider the process on the right-hand side of the previous inequality. Integrating with respect to s and applying the expectation operator we have

$$
(4.2) \qquad
\begin{aligned}
&\mathbb{E}\left[\int_0^T \left|\frac{\partial F}{\partial \xi}(t,\xi_{\bar{u}}(t))\right|\left(\sup_{\tau\in(0,T)}G(\tau)\right)\left(\sup_{\tau\in(0,T)}G'(\tau)\right)|\bar{u}(s) - u(s)|ds\right] \\
&= \|\bar{u} - u\|_{L^1}\,\mathbb{E}\left[\left|\frac{\partial F}{\partial \xi}(t,\xi_{\bar{u}}(t))\right|\left(\sup_{\tau\in(0,T)}G(\tau)\right)\left(\sup_{\tau\in(0,T)}G'(\tau)\right)\right].
\end{aligned}
$$

Now let us fix $\widetilde{p} \in \left(1, \min\left\{2, \frac{p}{\alpha}\right\}\right)$, where α is defined in hypothesis $(H4)$, and $\widetilde{q}$ such that $\frac{1}{\widetilde{p}} + \frac{1}{\widetilde{q}} = 1$. By Hölder's inequality we have

$$
\begin{aligned}
&\mathbb{E}\left[\left|\frac{\partial F}{\partial \xi}(t,\xi_{\bar{u}}(t))\right|\left(\sup_{\tau\in(0,T)}G(\tau)\right)\left(\sup_{\tau\in(0,T)}G'(\tau)\right)\right] \\
&\leq \mathbb{E}\left[\left|\frac{\partial F}{\partial \xi}(t,\xi_{\bar{u}}(t))\right|^{\widetilde{p}}\right]^{\frac{1}{\widetilde{p}}}\mathbb{E}\left[\left(\sup_{\tau\in(0,T)}G(\tau)\right)^{\widetilde{q}}\left(\sup_{\tau\in(0,T)}G'(\tau)\right)^{\widetilde{q}}\right]^{\frac{1}{\widetilde{q}}} \\
&\leq CL(t)\,\mathbb{E}\left[(1 + |\xi_{\bar{u}}(t)|^\alpha)^{\widetilde{p}}\right]^{\frac{1}{\widetilde{p}}} \\
&\leq CL(t)2^{1-\frac{1}{\widetilde{p}}}(1 + \mathbb{E}[|\xi_{\bar{u}}(t)|^{\alpha\widetilde{p}}]^{\frac{1}{\widetilde{p}}}) \\
&\leq CL(t)2^{1-\frac{1}{\widetilde{p}}}(1 + \mathbb{E}[|\xi_{\bar{u}}(t)|^p]^{\frac{\alpha}{p}}),
\end{aligned}
$$

where we used hypothesis $(H4)$, Lemma 2.8 and Hölder's inequality a second time with exponent $\frac{p}{\alpha\widetilde{p}} > 1$. By Lemma 2.6 we conclude that

$$
\mathbb{E}\left[\left|\frac{\partial F}{\partial \xi}(t,\xi_{\bar{u}}(t))\right|\left(\sup_{\tau\in(0,T)}G(\tau)\right)\left(\sup_{\tau\in(0,T)}G'(\tau)\right)\right] \leq CL(t).
$$

Going back to Equation (4.2) we have

$$\mathbb{E}\left[\int_0^T \left|\frac{\partial F}{\partial \xi}(t, \xi_{\bar{u}}(t))e^{A(t)-A(s)}\frac{G(t)}{G(s)}(\bar{u}(s)-u(s))\right|ds\right] \leq C\left\|\bar{u}-u\right\|_{L^1}L(t)$$

and then, integrating with respect to t,

$$\int_0^T \mathbb{E}\left[\int_0^T \left|\frac{\partial F}{\partial \xi}(t, \xi_{\bar{u}}(t))e^{A(t)-A(s)}\frac{G(t)}{G(s)}(\bar{u}(s)-u(s))\right|ds\right]dt \leq C\left\|\bar{u}-u\right\|_{L^1}\left\|L\right\|_{L^1}.$$

Hence, we can use Fubini's theorem in Equation (4.1) to achieve

$$J[u]-J[\bar{u}] \geq \int_0^T (\bar{u}(s)-u(s))\int_s^T \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_{\bar{u}}(t))e^{A(t)-A(s)}\frac{G(t)}{G(s)}\right]dt\,ds = 0,$$

$\bar{u}$ being a solution of (3.1). The fact that $u \in L^1$ is arbitrary concludes the proof. $\qquad\square$

The previous result is strictly linked with the convexity hypothesis (*H5*). Indeed, such hypothesis actually implies the convexity of the operator J.

Proposition 4.2. *Suppose Hypotheses (H1) to (H5) are satisfied. Then J is convex. Moreover, if (H5+) is satisfied, J is strictly convex.*

Proof. Let us consider $u_1, u_2 \in L^1(0, T)$ and $\theta \in [0, 1]$. Then we have

$$\begin{aligned}
J[\theta u_1 + (1-\theta)u_2] &= \mathbb{E}\left[\int_0^T F(t, \mathcal{S}_0(z-(\theta u_1+(1-\theta)u_2))(t))dt\right] \\
&= \mathbb{E}\left[\int_0^T F(t, \theta\,\mathcal{S}_0(z-u_1)(t)+(1-\theta)\,\mathcal{S}_0(z-u_2)(t))dt\right] \\
&\leq \theta\,\mathbb{E}\left[\int_0^T F(t, \mathcal{S}_0(z-u_1)(t))dt\right] + (1-\theta)\,\mathbb{E}\left[\int_0^T F(t, \mathcal{S}_0(z-u_2)(t))dt\right] \\
&= \theta J[u_1] + (1-\theta)J[u_2],
\end{aligned}$$

where we used hypothesis (*H5*) and the third statement of Corollary 2.3. This proves that J is convex. Now let us suppose $u_1 \neq u_2$ (that is to say there exists a set $E \subseteq (0, T)$ with $|E| > 0$ and $u_1 \neq u_2$ on E), $\theta \in (0, 1)$ and (*H5+*) holds. By definition of the solution map, there exists $\overline{\Omega}$ such that

    I $\mathbb{P}(\overline{\Omega}) > 0$;

    II for any $\omega \in \overline{\Omega}$, $\mathcal{S}_0(z-u_i)(\cdot, \omega)$ is continuous in $[0, T]$ for $i = 1, 2$;

    III for any $\omega \in \overline{\Omega}$, there exists $t(\omega)$ such that $\mathcal{S}_0(z-u_1)(t(\omega), \omega) \neq \mathcal{S}_0(z-u_2)(t(\omega), \omega)$.

In particular, combining II and III we have that for any $\omega \in \overline{\Omega}$ there exists an interval $I(\omega)$ such that $\mathcal{S}_0(z-u_1)(t, \omega) \neq \mathcal{S}_0(z-u_2)(t, \omega)$ for any $t \in I(\omega)$. By using the third statement of Corollary 2.3 we get

$$\begin{aligned}
J[\theta u_1 + (1-\theta)u_2] &= \mathbb{E}\left[\int_0^T F(t, \mathcal{S}_0(z-(\theta u_1+(1-\theta)u_2))(t))dt\right] \\
&= \mathbb{E}\left[\int_0^T F(t, \theta\,\mathcal{S}_0(z-u_1)(t)+(1-\theta)\,\mathcal{S}_0(z-u_2)(t))dt\right] \\
&= \mathbb{E}\left[\int_{I(\cdot)} F(t, \theta\,\mathcal{S}_0(z-u_1)(t)+(1-\theta)\,\mathcal{S}_0(z-u_2)(t))dt; \overline{\Omega}\right] \\
&\quad + \mathbb{E}\left[\int_{[0,T]\setminus I(\cdot)} F(t, \theta\,\mathcal{S}_0(z-u_1)(t)+(1-\theta)\,\mathcal{S}_0(z-u_2)(t))dt; \overline{\Omega}\right] \\
&\quad + \mathbb{E}\left[\int_0^T F(t, \theta\,\mathcal{S}_0(z-u_1)(t)+(1-\theta)\,\mathcal{S}_0(z-u_2)(t))dt; \Omega\setminus\overline{\Omega}\right],
\end{aligned}$$

where for any random variable Z and any event $B \in \mathcal{F}$ we use the notation $\mathbb{E}[Z; B] = \mathbb{E}[Z\mathbf{1}_B]$. Now let us observe that by Hypothesis $(H5+)$ we have

$$F(t, \theta\, \mathcal{S}_0(z - u_1)(t, \omega) + (1 - \theta)\, \mathcal{S}_0(z - u_2)(t, \omega))$$
$$\leq \theta F(t, \mathcal{S}_0(z - u_1)(t, \omega)) + (1 - \theta)F(t, \mathcal{S}_0(z - u_2)(t, \omega)), \ \forall \omega \in \Omega, \ \forall t \in [0, T],$$

and the inequality is strict for any $\omega \in \overline{\Omega}$ and $t \in I(\omega)$. Thus we get

$$J[\theta u_1 + (1 - \theta)u_2] < \theta\, \mathbb{E}\left[\int_{I(\cdot)} F(t, \mathcal{S}_0(z - u_1)(t))dt; \overline{\Omega}\right] + (1 - \theta)\, \mathbb{E}\left[\int_{I(\cdot)} F(t, \mathcal{S}_0(z - u_2)(t))dt; \overline{\Omega}\right]$$

$$+ \theta\, \mathbb{E}\left[\int_{[0,T]\setminus I(\cdot)} F(t, \mathcal{S}_0(z - u_1)(t))dt; \overline{\Omega}\right] + (1 - \theta)\, \mathbb{E}\left[\int_{[0,T]\setminus I(\cdot)} F(t, \mathcal{S}_0(z - u_2)(t))dt; \overline{\Omega}\right]$$

$$+ \theta\, \mathbb{E}\left[\int_0^T F(t, \mathcal{S}_0(z - u_1)(t))dt; \Omega \setminus \overline{\Omega}\right] + (1 - \theta)\, \mathbb{E}\left[\int_0^T F(t, \mathcal{S}_0(z - u_2)(t))dt; \Omega \setminus \overline{\Omega}\right]$$

$$= \theta J[u_1] + (1 - \theta)J[u_2],$$

concluding the proof. $\square$

Remark 4.3. Let us observe that Theorem 4.1 can be seen as a direct consequence of Proposition 4.2 and Remark 3.3, by using the inequality

$$J[u] - J[\bar{u}] \geq \partial_{\bar{u}} J[u - \bar{u}],$$

implied by the convexity of $J[u]$.

Another direct consequence of Proposition 4.2 is given by the following Corollary.

Corollary 4.4. *Suppose Hypotheses (H1) to (H5+) are satisfied. Then Equation* (3.1) *admits at most one solution.*

Proof. Let us suppose $\bar{u}_1$ and $\bar{u}_2$ are two solutions of Equation (3.1). Then, by Theorem 4.1 we know that both $\bar{u}_1$ and $\bar{u}_2$ are global minimizers of J. However, Proposition 4.2 tells us that J is strictly convex, hence it admits a unique global minimizer and then $\bar{u}_1 = \bar{u}_2$. $\square$

5. MINIMIZING FAMILIES

Up to now we are not able to show that J is coercive, which should be the main ingredient, together with lower semicontinuity, to prove the existence of a minimizer. This is due to the fact that, since $\xi_u(t) = \mathcal{S}_0(z - u)(t)$ depends on a sort of primitive function of u, classical lower bounds such as $F(t, \xi) \geq L(1 + |\xi|^p)$ are not enough to guarantee coercivity. For this reason, we focus instead on exploiting some minimizing families for J, i.e. a family of functions $\{u_\delta\}_{\delta > 0}$ with the property that, for any $\varepsilon > 0$, there exists $\delta_0 > 0$ such that if $\delta \in (0, \delta_0)$ it holds

$$m \leq J[u_\delta] \leq m + \varepsilon$$

where $m = \inf_{u \in L^1(0,T)} J[u]$.

First of all, we observe that J is a continuous functional on $L^1(0, T)$.

Proposition 5.1. *Let hypotheses (H1) to (H4) hold. Then $J : L^1(0, T) \to \mathbb{R}$ is continuous, i.e. for any fixed $u_1 \in L^1(0, T)$, for any $\varepsilon > 0$ there exists $r > 0$ (possibly depending on u_1) such that*

$$\forall u_2 \in B_r(u_1), \ |J[u_2] - J[u_1]| < \varepsilon,$$

where $B_r(u_1) = \{u_2 \in L^1(0, T) : \ \|u_2 - u_1\|_{L^1(0,T)} < r\}$.

Proof. Fix $\varepsilon > 0$ and let $u_1, u_2 \in L^1(0, T)$ with $\|u_1 - u_2\|_{L^1(0,T)} < r$, where r will be defined in what follows. Then

$$(5.1) \quad |J[u_2] - J[u_1]| \leq \mathbb{E}\left[\int_0^T |F(t, \xi_{u_2}(t)) - F(t, \xi_{u_1}(t))|dt\right]$$

$$\leq \mathbb{E}\left[\int_0^T \int_0^1 \left|\frac{\partial F}{\partial \xi}(t, \theta\xi_{u_2}(t) + (1 - \theta)\xi_{u_1}(t))\right| |\xi_{u_2}(t) - \xi_{u_1}(t)|dt\right],$$

where we used hypothesis $(H3)$. Let us consider p as in $(H1)$ and α as in $(H4)$ and, let $\widetilde{p} \in \left(1, \min\left\{2, \frac{p}{\alpha}\right\}\right)$ and $\widetilde{q}$ such that $\frac{1}{\widetilde{p}} + \frac{1}{\widetilde{q}} = 1$. By Hölder's inequality we have

$$(5.2) \quad \mathbb{E}\left[\left|\frac{\partial F}{\partial \xi}(t, \theta\xi_{u_2}(t) + (1-\theta)\xi_{u_1}(t))\right| |\xi_{u_2}(t) - \xi_{u_1}(t)|\right]$$

$$\leq \mathbb{E}\left[\left|\frac{\partial F}{\partial \xi}(t, \theta\xi_{u_2}(t) + (1-\theta)\xi_{u_1}(t))\right|^{\widetilde{p}}\right]^{\frac{1}{\widetilde{p}}} \mathbb{E}[|\xi_{u_2}(t) - \xi_{u_1}(t)|^{\widetilde{q}}]^{\frac{1}{\widetilde{q}}}.$$

Concerning the first term, we have, by hypothesis $(H4)$

$$\mathbb{E}\left[\left|\frac{\partial F}{\partial \xi}(t, \theta\xi_{u_2}(t) + (1-\theta)\xi_{u_1}(t))\right|^{\widetilde{p}}\right]^{\frac{1}{\widetilde{p}}} \leq L(t)\,\mathbb{E}[(1 + |\theta\xi_{u_2}(t) + (1-\theta)\xi_{u_1}(t)|^\alpha)^{\widetilde{p}}]^{\frac{1}{\widetilde{p}}}$$

$$\leq 2^{1-\frac{1}{\widetilde{p}}} L(t)(1 + \mathbb{E}[|\theta\xi_{u_2}(t) + (1-\theta)\xi_{u_1}(t)|^{\alpha\widetilde{p}}]^{\frac{1}{\widetilde{p}}})$$

$$\leq 2^{1-\frac{1}{\widetilde{p}}} L(t)(1 + \mathbb{E}[|\theta\xi_{u_2}(t) + (1-\theta)\xi_{u_1}(t)|^p]^{\frac{\alpha}{p}}),$$

where we used again Hölder's inequality with exponent $\frac{p}{\alpha\widetilde{p}} > 1$. Next, since $\xi_{u_i} = \mathcal{S}_0(z - u_i)$, by Corollary 2.3 and Lemma 2.4 we get

$$\mathbb{E}\left[\left|\frac{\partial F}{\partial \xi}(t, \theta\xi_{u_2}(t) + (1-\theta)\xi_{u_1}(t))\right|^{\widetilde{p}}\right]^{\frac{1}{\widetilde{p}}} \leq 2^{1-\frac{1}{\widetilde{p}}} L(t)(1 + \mathbb{E}[|S_0(z - \theta u_2 - (1-\theta)u_1)|^p]^{\frac{\alpha}{p}})$$

$$\leq CL(t)\left(1 + \left(\int_0^T \mathbb{E}[|z(t) - (\theta u_2 + (1-\theta)u_1)|^p]^{\frac{1}{p}}\right)^\alpha\right)$$

$$\leq CL(t)\left(1 + 2^{\alpha-\frac{\alpha}{p}}\left(\int_0^T \mathbb{E}[|z(t)|^p]^{\frac{1}{p}}dt + \int_0^T |\theta u_2(t) + (1-\theta)u_1(t)|dt\right)^\alpha\right)$$

$$\leq CL(t)\left(1 + 2^{2\alpha-\frac{\alpha}{p}-1}\left(\left(\int_0^T \mathbb{E}[|z(t)|^p]^{\frac{1}{p}}dt\right)^\alpha + \left(\int_0^T |\theta u_2(t) + (1-\theta)u_1(t)|dt\right)^\alpha\right)\right)$$

$$\leq CL(t)\left(1 + \left(\int_0^T |\theta u_2(t) + (1-\theta)u_1(t)|dt\right)^\alpha\right)$$

$$\leq CL(t)\left(1 + \theta\|u_2\|_{L^1(0,T)}^\alpha + (1-\theta)\|u_1\|_{L^1(0,T)}^\alpha\right)$$

where C is independent of u_1 and u_2, and we used the fact that $\left(\int_0^T \mathbb{E}[|z(t)|^p]^{\frac{1}{p}}dt\right)^\alpha < +\infty$ and that $t \geq 0 \mapsto t^\alpha$ is a convex function. We can assume, without loss of generality, that $r \leq 1$. Since $\|u_2 - u_1\|_{L^1(0,T)} < r \leq 1$, we achieve

$$(5.3) \quad \begin{aligned} &\mathbb{E}\left[\left|\frac{\partial F}{\partial \xi}(t, \theta\xi_{u_2}(t) + (1-\theta)\xi_{u_1}(t))\right|^{\widetilde{p}}\right]^{\frac{1}{\widetilde{p}}} \\ &\leq CL(t)\left(1 + \theta 2^{\alpha-1}\|u_2 - u_1\|_{L^1(0,T)}^\alpha + (1 - \theta + 2^{\alpha-1})\|u_1\|_{L^1(0,T)}^\alpha\right) \\ &< CL(t)\left(1 + \|u_1\|_{L^1(0,T)}^\alpha\right). \end{aligned}$$

On the other hand, still recalling that $\xi_{u_i} = \mathcal{S}_0(z - u_i)$ and by Corollary 2.3 and Lemma 2.6, we get

$$(5.4) \qquad \mathbb{E}[|\xi_{u_2}(t) - \xi_{u_1}(t)|^{\widetilde{q}}]^{\frac{1}{\widetilde{q}}} = \mathbb{E}[|\mathcal{S}_0(u_2 - u_1)(t)|^{\widetilde{q}}]^{\frac{1}{\widetilde{q}}} \leq C\|u_2 - u_1\|_{L^1(0,T)} < Cr.$$

Combining Equations (5.3) and (5.4) in Equation (5.2) we get

$$\mathbb{E}\left[\left|\frac{\partial F}{\partial \xi}(t, \theta\xi_{u_2}(t) + (1-\theta)\xi_{u_1}(t))\right| |\xi_{u_2}(t) - \xi_{u_1}(t)|\right] < CL(t)\left(1 + \|u_1\|_{L^1(0,T)}^\alpha\right)r.$$

Using the previous inequality in (5.1) together with Fubini's theorem, since the integrand is non-negative, we know there exists a constant $\overline{C} > 0$, independent of u_1 and u_2, such that

$$|J[u_2] - J[u_1]| < \overline{C}\left(1 + \|u_1\|_{L^1(0,T)}^\alpha\right) r,$$

Setting

$$r = \min\left\{1, \frac{\varepsilon}{\overline{C}\left(1 + \|u_1\|_{L^1(0,T)}^\alpha\right)}\right\}$$

we conclude the proof. □

Remark 5.2. If, moreover, Hypothesis (*H5*) holds, then J is also weakly lower semicontinuous in L^1. This is a direct consequence of Mazur's theorem [15, Theorem 3.9].

Let us introduce the set of simple functions:

$$\mathfrak{S} = \left\{u \in L^1(0,T): \ \exists N \in \mathbb{N}, \{B_i\}_{i \leq N} \subseteq \mathcal{B}([0,T]), \ \{b_i\}_{i \leq N} \subseteq \mathbb{R}, \ u = \sum_{i=1}^N b_i \mathbf{1}_{B_i}\right\},$$

where $\mathcal{B}([0,T])$ is the Borel σ-algebra on the interval $[0,T]$. As a direct consequence of Proposition 5.1 we have the following Corollary.

Corollary 5.3. *Let hypotheses (H1) to (H4) hold. Then*

$$\inf_{u \in L^1(0,T)} J[u] = \inf_{u \in \mathfrak{S}} J[u].$$

Proof. It follows from the fact that $J : L^1(0,T) \to \mathbb{R}$ is continuous (by Propostion 5.1) and the fact that $\mathfrak{S}$ is dense in $L^1(0,T)$ (see, for instance, [39, Theorem 3.13]). □

Now we want to *penalize* our functional J to obtain a coercive functional. To do this, let us first give the following definition.

Definition 5.1. *A function* $\Psi : [0,+\infty) \to [0,+\infty)$ *is called a Young function (see [37]) if there exists a function* $\psi : [0,+\infty) \to [0,+\infty)$ *such that*

$$\Psi(t) = \int_0^t \psi(s)ds, \ t \geq 0$$

and ψ *satisfies the following properties:*
- $\psi(0) = 0$;
- $\psi(s) > 0$ *for any* $s > 0$;
- ψ *is right-continuous;*
- ψ *is non-decreasing;*
- $\lim_{s \to +\infty} \psi(s) = +\infty$.

Young functions satisfy different important properties. Here we recall some of them (see [37, Lemma 4.2.2]).

Lemma 5.4. *Any Young function* Ψ *is continuous, non-negative, strictly increasing and convex. Moreover it holds* $\Psi(0) = 0$, $\lim_{t \to 0^+} t^{-1}\Psi(t) = 0$ *and* $\lim_{t \to +\infty} t^{-1}\Psi(t) = +\infty$.

Remark 5.5. Let us observe that if Ψ is a Young function then the function $u \in \mathbb{R} \mapsto \Psi(|u|)$ is differentiable. Indeed this is clearly true for $u \neq 0$ with derivative

$$\frac{d\Psi(|u|)}{du} = \frac{u}{|u|}\psi(|u|), \quad u \neq 0.$$

Observing that $-1 \leq \frac{u}{|u|} \leq 1$ and that, ψ being right-continuous with $\psi(0) = 0$, it holds $\lim_{u \to 0} \frac{d\Psi(|u|)}{du} = 0$ and then $\Psi(|u|)$ is differentiable at 0 with derivative 0. We will use the notation $\frac{u}{|u|}\psi(|u|)$ for any $u \in \mathbb{R}$, implying the 0 value as $u = 0$.

Fix any Young function Ψ, $\delta > 0$ and define the following functional

$$J_{\delta,\Psi} : L^1(0,T) \to J[u] + \delta\,\mathcal{F}_\Psi[u]$$

where

$$\mathcal{F}_\Psi[u] = \int_0^T \Psi(|u(t)|)dt.$$

Remark 5.6. There exist some $u \in L^1(0,T)$ such that $J_{\delta,\Psi}[u] = +\infty$ (see, for instance, [37, Remark 4.2.4]). In particular we can define the Orlicz class

$$\mathcal{L}^\Psi(0,T) = \left\{ u \in L^1(0,T) : \int_0^T \Psi(|u(t)|)dt < +\infty \right\}$$

and observe that $J_{\delta,\Psi}[u] < +\infty$ if and only if $u \in \mathcal{L}^\Psi(0,T)$. Let us stress out that $\mathcal{L}^\Psi(0,T)$ is in general not a vector space. Actually, $\mathcal{L}^\Psi(0,T)$ is a vector space if and only if $\Psi \in \Delta_2$, i.e. there exists a constant $k > 0$ such that $\Psi(2t) \leq k\Psi(t)$ for any $t \geq 0$ (see [37, Theorem 4.5.3]). In particular, if $\Psi \in \Delta_2$, we have that $\mathcal{L}^\Psi(0,T) = L^\Psi(0,T)$, where $L^\Psi(0,T)$ is defined as

$$L^\Psi(0,T) = \left\{ u \in L^1(0,T) : \exists \lambda > 0 : \int_0^T \Psi\left(\frac{|u(t)|}{\lambda}\right) dt < +\infty \right\}$$

and is a Banach space when equipped with the norm

$$\|u\|_{L^\Psi(0,T)} := \inf\left\{ \lambda > 0 : \int_0^T \Psi\left(\frac{|u(t)|}{\lambda}\right) dt \leq 1 \right\}.$$

See [37, Chapter 4] for further details.

First of all, we want to show that for any $\delta > 0$ and any Young function Ψ, the functional $J_{\delta,\Psi}$ admits a minimum. To do this, we need the following preliminary result.

Proposition 5.7. *For any Young function $\Psi \in \Delta_2$, the functional $\mathcal{F}_\Psi : L^1(0,T) \mapsto \mathbb{R} \cup \{+\infty\}$ is weakly lower semicontinuous.*

The previous result relies on classical arguments in Calculus of Variation (see [2, 16, 42]). We provide its proof in Appendix B for completeness.

Now we are ready to show that $J_{\delta,\Psi}$ admits a minimum.

Theorem 5.8. *Let $\delta > 0$ and $\Psi \in \Delta_2$ be a Young function. Suppose hypotheses (H1) to (H5) hold. Then there exists a function $\bar{u}_{\delta,\Psi} \in L^1(0,T)$ such that*

$$\min_{u \in L^1(0,T)} J_{\delta,\Psi}[u] = J_{\delta,\Psi}[\bar{u}_{\delta,\Psi}].$$

Moreover, if hypothesis (H5+) holds or if Ψ is strictly convex, then $\bar{u}_{\delta,\Psi} \in L^1(0,T)$ is unique.

Proof. Let us first observe that $J_{\delta,\Psi}$ is weakly lower semicontinuous. To do this, observe that $J_{\delta,\Psi}$ is the sum of two convex functionals J and $\delta\,\mathcal{F}_\Psi$, thus it is convex. Moreover, J is continuous by Proposition 5.1 and then, in particular, lower semicontinuous, while $\delta\,\mathcal{F}_\Psi$ is lower semicontinuous by Proposition 5.7. Thus $J_{\delta,\Psi}$ is convex and lower semicontinuous and then weakly lower semicontinuous by a direct application of Mazur's theorem [15, Theorem 3.9].

Now let us consider a minimizing sequence $\{u_n\}_{n \in \mathbb{N}} \subset L^1(0,T)$ of $J_{\delta,\Psi}$, i.e. $\{u_n\}_{n \in \mathbb{N}}$ is such that $J_{\delta,\Psi}[u_n] \downarrow \inf_{u \in L^1(0,T)} J_{\delta,\Psi}[u]$. Let us consider any function $f \in L^\Psi(0,T)$. Then $J_{\delta,\Psi}[f] < +\infty$ by definition of L^Ψ. In particular this implies that $\inf_{u \in L^1(0,T)} J_{\delta,\Psi}[u] < +\infty$ and we can suppose $J_{\delta,\Psi}[u_1] < +\infty$. Let us observe that

$$\delta\,\mathcal{F}_\Psi[u_n] \leq J_{\delta,\Psi}[u_n] \leq J_{\delta,\Psi}[u_1], \ \forall n \in \mathbb{N}$$

and then there exists a constant $C(\delta) = \frac{J_{\delta,\Psi}[u_1]}{\delta}$ such that

$$\mathcal{F}_\Psi[u_n] \leq C(\delta), \ \forall n \in \mathbb{N}.$$

By the de la Vallée-Poussin theorem (see [30, Theorem T22]), we know that the sequence u_n is uniformly integrable. By the Dunford-Pettis theorem (see [13, Theorem 4.30]) we have that the sequence $\{u_n\}$ is

weakly relatively compact in $L^1(0,T)$ and then there exists $\overline{u}_{\delta,\Psi} \in L^1(0,T)$ such that $u_n \rightharpoonup \overline{u}_{\delta,\Psi}$. By weak semicontinuity of $J_{\delta,\Psi}$ we have

$$\inf_{u \in L^1(0,T)} J_{\delta,\Psi}[u] = \lim_{n \to +\infty} J_{\delta,\Psi}[u_n] \geq J_{\delta,\Psi}[\overline{u}_{\delta,\Psi}] \geq \inf_{u \in L^1(0,T)} J_{\delta,\Psi}[u],$$

concluding the proof of the first statement. Concerning the second statement, it follows from the fact that if hypothesis (H5+) holds or if Ψ is strictly convex, then $J_{\delta,\Psi}$ is a strictly convex functional and then the minimum is unique. $\qquad\square$

Remark 5.9. Let us first observe that for any $\delta > 0$ it holds $\overline{u}_{\delta,\Psi} \in L^\Psi(0,T)$.
Moreover, we can prove that the functional $J_{\delta,\Psi}$ is coercive with respect to the weak topology in $L^1(0,T)$, i.e. for any $M > 0$ there exists a weakly compact set $K_M \subset L^1(0,T)$ such that $u \in L^1(0,T) \setminus K_M$ implies $J_{\delta,\Psi}[u] > M$. Precisely, if $J_{\delta,\Psi}[u] \leq M$, we have

$$\delta \, \mathcal{F}_\Psi[u] \leq J_{\delta,\Psi}[u] \leq M$$

and then $\mathcal{F}_\Psi[u] \leq M/\delta$. The set $U_M = \{u \in L^1(0,T) : \ \mathcal{F}_\Psi[u] \leq M/\delta\}$ is uniformly integrable by the de la Vallée-Poussin theorem and then it is weakly relatively compact by the Dunford-Pettis theorem. Let $K_M = \overline{U_M}$, where the closure is taken in the weak topology of $L^1(0,T)$, so that K_M is weakly compact. Then $J_{\delta,\Psi}[u] \leq M$ implies $u \in K_M$ and, by contrapositive, we have that $u \in L^1(0,T) \setminus K_M$ implies $J_{\delta,\Psi}[u] > M$.

Now we want to show that the penalization procedure (i.e. defining the functional $J_{\delta,\Psi}$ as J plus a penalization term $\delta \, \mathcal{F}_\Psi$) generates a minimizing family for J.

Theorem 5.10. *Let hypotheses (H1) to (H5) hold and consider a Young function $\Psi \in \Delta_2$. The family of functions $\{\overline{u}_{\delta,\Psi}\}_{\delta>0}$ defined in Theorem 5.8 constitute a minimizing family for J.*

Proof. Set $m = \inf_{u \in L^1(0,T)} J[u]$. Let us first observe that $\mathfrak{S} \subset L^\Psi(0,T)$. Indeed, if $u \in \mathfrak{S}$ then there exist $N \in \mathbb{N}$, $\{B_i\}_{i \leq N} \subset \mathcal{B}([0,T])$ and $\{b_i\}_{i \leq N} \subset \mathbb{R}$ such that $u = \sum_{i=1}^N b_i \mathbf{1}_{B_i}$. We can suppose, without loss of generality, that $\bigcup_{i \leq N} B_i = [0,T]$ and that $B_i \cap B_j = \emptyset$ for any $i \neq j$. Hence we have

$$\int_0^T \Psi(|u(t)|)dt = \sum_{i=1}^N \int_{B_i} \Psi(|b_i|)dt = \sum_{i=1}^N \Psi(|b_i|)|B_i| \leq T \max_{i \leq N} \Psi(|b_i|) < +\infty.$$

Now let us consider any $u \in \mathfrak{S}$ and observe that $J_{\delta,\Psi}[u] < +\infty$. Moreover, we have

$$m \leq J[\overline{u}_{\delta,\Psi}] \leq J_{\delta,\Psi}[\overline{u}_{\delta,\Psi}] \leq J_{\delta,\Psi}[u] = J[u] + \delta \, \mathcal{F}_\Psi[u].$$

Taking the limit superior and inferior as $\delta \to 0$ we get

$$m \leq \liminf_{\delta \to 0} J[\overline{u}_{\delta,\Psi}] \leq \limsup_{\delta \to 0} J[\overline{u}_{\delta,\Psi}] \leq J[u].$$

Being $u \in \mathfrak{S}$ arbitrary, we can take the infimum on $\mathfrak{S}$ and use Corollary 5.3 to achieve

$$m \leq \liminf_{\delta \to 0} J[\overline{u}_{\delta,\Psi}] \leq \limsup_{\delta \to 0} J[\overline{u}_{\delta,\Psi}] \leq m,$$

obtaining $\lim_{\delta \to 0} J[\overline{u}_{\delta,\Psi}] = m$ and concluding the proof. $\qquad\square$

Last theorem provides a theoretical way to construct a minimizing family for the functional J. Let us first stress out that, under an additional regularity assumption, the previous approach actually proves the existence of a minimizer for J.

Theorem 5.11. *Let hypotheses (H1) to (H5) hold and suppose there exist two Young functions $\Psi_i \in \Delta_2$, $i = 1, 2$, and two constants $C, \delta_0 > 0$ such that*

$$(5.5) \qquad\qquad \mathcal{F}_{\Psi_2}[\overline{u}_{\delta,\Psi_1}] \leq C$$

for any $\delta \in (0,\delta_0)$, where the functions $\overline{u}_{\delta,\Psi_1}$ are defined in Theorem 5.8. Then there exists $\overline{u} \in L^1(0,T)$ such

$$\inf_{u \in L^1(0,T)} J[u] = J[\overline{u}]$$

and $\overline{u}_{\delta,\Psi_1} \rightharpoonup \overline{u}$ as $\delta \to 0$.
Moreover, if hypothesis (H5+) holds, then $\overline{u}$ is unique.

Proof. Let us consider any sequence $\delta_n \downarrow 0$ with $\delta_1 < \delta_0$. Equation (5.5) implies, via the de la Vallée-Poussin theorem, that the sequence $\{\overline{u}_{\delta_n,\Psi_1}\}_{n \in \mathbb{N}}$ is uniformly integrable and then the Dunford-Pettis theorem ensures that it is weakly relatively compact. Hence, there exists $\overline{u}$ such that $\overline{u}_{\delta_n,\Psi_1} \rightharpoonup \overline{u}$.

By Theorem 5.10 we know that $\lim_n J[\overline{u}_{\delta_n,\Psi_1}] = \inf_{u \in L^1(0,T)} J[u] =: m$. On the other hand, being J convex and continuous in $L^1(0,T)$, we know that it is weakly lower semicontinuous and then

$$m = \lim_n J[\overline{u}_{\delta_n,\Psi_1}] \geq J[\overline{u}] \geq m$$

thus $J[\overline{u}] = m$, concluding the first part of the proof. The second statement follows from the strict convexity of J. $\qquad\square$

Remark 5.12. The previous theorem can be proved directly by using the fact that $J_{\delta_n,\Psi}$ is a monotone sequence of operators pointwise converging to J, that is convex and lower semicontinuous, thus it also Γ-converges towards J in the weak topology of $L^1(0,T)$ (see [12, Remark 1.40]). Let us also recall that, being $L^\infty(0,T)$ not separable, the weak topology of $L^1(0,T)$ is not metrizable on closed balls, hence the more general definition of Γ-convergence on topological spaces has to be considered.

Let us stress out that if a minimizer of J exists *in the right space*, then we are under the hypotheses of the previous theorem.

Corollary 5.13. *Let hypotheses (H1) to (H5) hold and consider a Young function $\Psi \in \Delta_2$. Suppose there exists $\overline{u} \in L^\Psi(0,T)$ such that*

$$J[\overline{u}] = \min_{u \in L^1(0,T)} J[u].$$

Then, the family $\{\overline{u}_{\delta,\Psi}\}_{\delta>0}$ defined in Theorem 5.8 satisfies the hypotheses of Theorem 5.11 with $\Psi_1 = \Psi_2 = \Psi$. Moreover, if hypothesis (H5+) holds, then $\overline{u}_{\delta_n,\Psi} \rightharpoonup \overline{u}$ for some sequence $\delta_n \to 0$.

Proof. We have to show that there exists a constant $C > 0$ such that $\mathcal{F}_\Psi[\overline{u}_{\delta,\Psi}] \leq C$ for any $\delta > 0$. To do this, let us recall that $\overline{u}_{\delta,\Psi}$ is a global minimizer of $J_{\delta,\Psi}$, thus we have

$$J[\overline{u}_{\delta,\Psi}] + \delta \, \mathcal{F}_\Psi[\overline{u}_{\delta,\Psi}] = J_{\delta,\Psi}[\overline{u}_{\delta,\Psi}] \leq J_{\delta,\Psi}[\overline{u}] = J[\overline{u}] + \delta \, \mathcal{F}_\Psi[\overline{u}].$$

On the other hand, being $\overline{u}$ the global minimizer of J, we have

$$J[\overline{u}_{\delta,\Psi}] + \delta \, \mathcal{F}_\Psi[\overline{u}_{\delta,\Psi}] \leq J[\overline{u}] + \delta \, \mathcal{F}_\Psi[\overline{u}] \leq J[\overline{u}_{\delta,\Psi}] + \delta \, \mathcal{F}_\Psi[\overline{u}],$$

that is to say

$$\mathcal{F}_\Psi[\overline{u}_{\delta,\Psi}] \leq \mathcal{F}_\Psi[\overline{u}].$$

Setting $C = \mathcal{F}_\Psi[\overline{u}]$, since $C < +\infty$ by hypothesis, we conclude the proof. $\qquad\square$

Remark 5.14. Actually, de la Vallée-Poussin theorem tells us that if we have a global minimizer $\overline{u} \in L^1(0,T)$ for J, then there exists a Young function Ψ such that $\overline{u} \in \mathcal{L}^\Psi(0,T)$. In such case, the previous theorem holds by choosing Ψ as Young function even if $\Psi \notin \Delta_2$. In conclusion, if J admits a global minimizer $\overline{u} \in L^1(0,T)$, then there exists a Young function Ψ such that $\overline{u}_{\delta,\Psi}$ weakly converge towards a (possibly different) minimizer of J and, if $\overline{u}$ is unique, then $\overline{u}_{\delta,\Psi} \rightharpoonup \overline{u}$.

As a consequence of the weak convergence of the minimizers we obtain a form of weak convergence of the approximating processes.

Theorem 5.15. *Let hypotheses (H1) to (H5+) hold and suppose J admits a global minimizer $\overline{u} \in L^{\widetilde{p}}(0,T)$ for some $\widetilde{p} > 2\sqrt{2}$ and consider $\Psi(t) = t^{\widetilde{p}}$. Let $\{\overline{u}_{\delta,\Psi}\}_{\delta>0}$ be the family defined in Theorem 5.8. Let $X_0 \in L^2(0,T)$. Then there exists a sequence $\delta_n \to 0$ such that $\mathcal{S}_{X_0}\overline{u}_{\delta_n,\Psi} \Rightarrow \mathcal{S}_{X_0}\overline{u}$ in $C([0,T])$ in distribution.*

Proof. Let $\mathcal{A} = \{\omega \in \Omega : G(\cdot,\omega) \in C([0,T])\}$ and recall that $\mathbb{P}(\mathcal{A}) = 1$. Fix $t > 0$, $\omega \in \mathcal{A}$ and observe that, by Corollary 2.3 and Equation (2.2),

$$|\mathcal{S}_{X_0}\overline{u}_{\delta,\Psi}(t,\omega) - \mathcal{S}_{X_0}\overline{u}(t,\omega)| = |\mathcal{S}_0(\overline{u}_{\delta,\Psi} - \overline{u})(t,\omega)| = G(t,\omega)e^{A(t)}\left|\int_0^t \frac{e^{-A(s)}}{G(s,\omega)}\overline{u}_{\delta,\Psi}(s)ds - \int_0^t \frac{e^{-A(s)}}{G(s,\omega)}\overline{u}(s)ds\right|.$$

Consider $\delta_n \to 0$ as in Corollary 5.13, so that $\overline{u}_{\delta_n,\Psi} \rightharpoonup \overline{u}$. Since $\frac{e^{-A(\cdot)}}{G(\cdot,\omega)}$ is a continuous function, we have that

$$\int_0^t \frac{e^{-A(s)}}{G(s,\omega)}\overline{u}_{\delta_n,\Psi}(s)ds \to \int_0^t \frac{e^{-A(s)}}{G(s,\omega)}\overline{u}(s)ds$$

and then

$$\lim_{n\to+\infty} |\mathcal{S}_{X_0}\overline{u}_{\delta_n,\Psi}(t,\omega) - \mathcal{S}_{X_0}\overline{u}(t,\omega)| = 0, \ \forall\omega\in\mathcal{A}, \tag{5.6}$$

that is to say that, for fixed $t \in [0,T]$, it holds $\mathcal{S}_{X_0}\overline{u}_{\delta_n,\Psi}(t) \to \mathcal{S}_{X_0}\overline{u}(t)$ almost surely. Let us observe that this is enough to guarantee the convergence in any finite-dimensional distribution. Indeed, consider $N \in \mathbb{N}$ and $t_1,\ldots,t_N \in [0,T]$ and fix $\omega \in \mathcal{A}$. Since Equation (5.6) holds for any t_i, $i = 1,\ldots,N$, we have that for any fixed $\varepsilon > 0$ there exists $\nu_i \in \mathbb{N}$ such that if $n \geq \nu_i$ it holds

$$|\mathcal{S}_{X_0}\overline{u}_{\delta_n,\Psi}(t_i,\omega) - \mathcal{S}_{X_0}\overline{u}(t_i,\omega)| < \frac{\varepsilon}{N}.$$

Let $\nu = \max\{\nu_1,\ldots,\nu_N\}$ and consider $n \geq \nu$. By the triangular inequality we get

$$|(S_{X_0}\overline{u}_{\delta_n,\Psi}(t_i,\omega))_{i\leq N} - (\mathcal{S}_{X_0}\overline{u}(t_i,\omega))_{i\leq N}| \leq \sum_{i=1}^{N}|\mathcal{S}_{X_0}\overline{u}_{\delta_n,\Psi}(t_i,\omega) - \mathcal{S}_{X_0}\overline{u}(t_i,\omega))_{i\leq n}| < \varepsilon,$$

so that

$$\lim_{n\to+\infty}|(\mathcal{S}_{X_0}\overline{u}_{\delta_n,\Psi}(t_i,\omega))_{i\leq N} - (\mathcal{S}_{X_0}\overline{u}(t_i,\omega))_{i\leq N}| = 0, \ \forall\omega\in\mathcal{A}.$$

This implies that $(\mathcal{S}_{X_0}\overline{u}_{\delta_n,\Psi}(t_i))_{i\leq N} \to (\mathcal{S}_{X_0}\overline{u}(t_i))_{i\leq N}$ almost surely and thus in distribution. To extend the convergence in distribution to the whole paths, we need to show that the sequence $\mathcal{S}_{X_0}\overline{u}_{\delta,\Psi_n}$ is tight. Let us denote, for simplicity, $X_\delta = \mathcal{S}_{X_0}\overline{u}_{\delta,\Psi}$. Consider $0 \leq t_1 < t_2 \leq T$ and observe that, by Equation (2.2), it holds

$$X_\delta(t_2) - X_\delta(t_1) = (G(t_2)e^{A(t_2)} - G(t_1)e^{A(t_1)})\int_0^{t_2}\frac{e^{-A(s)}}{G(s)}\overline{u}_{\delta,\Psi}(s)ds + G(t_2)e^{A(t_1)}\int_{t_1}^{t_2}\frac{e^{-A(s)}}{G(s)}\overline{u}_{\delta,\Psi}(s)ds.$$

Being $\widetilde{p} > 2\sqrt{2}$, we have

$$\frac{4+\widetilde{p}}{3\widetilde{p}-1} < \frac{4+\widetilde{p}}{4} < \frac{4}{\widetilde{p}} < \frac{\widetilde{p}}{2} < \widetilde{p}-1.$$

Hence, we can consider $p_1 \in \left(\frac{4}{\widetilde{p}}, \frac{\widetilde{p}}{2}\right)$, so that

$$1 < \frac{p_1+1}{p_1} < 2 < \frac{\widetilde{p}}{p_1} < \frac{4\widetilde{p}}{4+\widetilde{p}},$$

and $\overline{p} \in \left(2, \min\left\{\frac{\widetilde{p}}{p_1}, 4\right\}\right)$. By convexity inequality,

$$\mathbb{E}[|X_\delta(t_2) - X_\delta(t_1)|^{\overline{p}}] \leq 2^{\overline{p}-1}\left(\mathbb{E}\left[|G(t_2)e^{A(t_2)} - G(t_1)e^{A(t_1)}|^{\overline{p}}\left|\int_0^{t_2}\frac{e^{-A(s)}}{G(s)}\overline{u}_{\delta,\Psi}(s)ds\right|^{\overline{p}}\right]\right.$$
$$+ \mathbb{E}\left[G^{\overline{p}}(t_1)e^{\overline{p}A(t_1)}\left|\int_{t_1}^{t_2}\frac{e^{-A(s)}}{G(s)}\overline{u}_{\delta,\Psi}(s)ds\right|^{\overline{p}}\right]\right) \tag{5.7}$$
$$= 2^{\overline{p}-1}(I_1 + I_2).$$

Let us first work with I_2. Using Hölder's inequality with exponent p_1 (and q_1 such that $1/p_1 + 1/q_1 = 1$) and Jensen's inequality we achieve

$$I_2 \leq \mathbb{E}[G^{\overline{p}q_1}(t_1)e^{\overline{p}q_1 A(t_1)}]^{\frac{1}{q_1}}(t_2-t_1)^{\overline{p}}\mathbb{E}\left[\left|\frac{1}{(t_2-t_1)}\int_{t_1}^{t_2}\frac{e^{-A(s)}}{G(s)}\overline{u}_{\delta,\Psi}(s)ds\right|^{\overline{p}p_1}\right]^{\frac{1}{p_1}}$$
$$\leq \mathbb{E}[G^{\overline{p}q_1}(t_1)e^{\overline{p}q_1 A(t_1)}]^{\frac{1}{q_1}}(t_2-t_1)^{\overline{p}-\frac{1}{p_1}}\mathbb{E}\left[\int_{t_1}^{t_2}\frac{e^{-\overline{p}p_1 A(s)}}{G^{\overline{p}p_1}(s)}|\overline{u}_{\delta,\Psi}(s)|^{\overline{p}p_1}ds\right]^{\frac{1}{p_1}}. \tag{5.8}$$

Set $\gamma_1 = \overline{p} - \frac{1}{p_1}$ with $\gamma_1 > 1$ by the choice of $\overline{p}$ and p_1. Now observe that

$$\mathbb{E}[G^{\overline{p}q_1}(t_1)e^{\overline{p}q_1 A(t_1)}]^{\frac{1}{q_1}} \leq \left(\sup_{t\in[0,T]} e^{\overline{p}A(t)} \right) \mathbb{E}\left[\sup_{t\in[0,T]} |G(t)|^{\overline{p}q_1} \right]^{\frac{1}{q_1}}$$

$$\leq \left(\sup_{t\in[0,T]} e^{\overline{p}A(t)} \right) \left(\frac{\overline{p}q_1}{\overline{p}q_1 - 1} \right)^{\overline{p}} e^{\frac{\overline{p}(\overline{p}q_1-1)}{2}T} =: C_1(T,\overline{p},p_1),$$

where we also used Doob's maximal inequality. On the other hand, we have

(5.9)

$$\mathbb{E}\left[\int_{t_1}^{t_2} \frac{e^{-\overline{p}p_1 A(s)}}{G^{\overline{p}p_1}(s)} |\overline{u}_{\delta,\Psi}(s)|^{\overline{p}p_1} ds \right]^{\frac{1}{p_1}} \leq \mathbb{E}\left[\int_0^T e^{\overline{p}p_1(s-A(s))} (G'(s))^{\overline{p}p_1} |\overline{u}_{\delta,\Psi}(s)|^{\overline{p}p_1} ds \right]^{\frac{1}{p_1}}$$

$$\leq \left(\sup_{t\in[0,T]} e^{\overline{p}(t-A(t))} \right) \left(\int_0^T |\overline{u}_{\delta,\Psi}|^{\overline{p}p_1}(s)ds \right)^{\frac{1}{p_1}} \mathbb{E}\left[\sup_{t\in[0,T]} |G'(t)|^{\overline{p}p_1} \right]^{\frac{1}{p_1}}$$

$$\leq \left(\sup_{t\in[0,T]} e^{\overline{p}(t-A(t))} \right) e^{\frac{\overline{p}(\overline{p}p_1-1)}{2}T} \left(\int_0^T |\overline{u}_{\delta,\Psi}|^{\overline{p}p_1}(s)ds \right)^{\frac{1}{p_1}}.$$

Concerning the last remaining integral in the previous inequality, let us observe that, by definition of $\overline{p}$ and p_1, it holds $\frac{\widetilde{p}}{\overline{p}p_1} > 1$ hence, we can use it as exponent in Hölder's inequality, obtaining

$$\left(\int_0^T |\overline{u}_{\delta,\Psi}|^{\overline{p}p_1}(s)ds \right)^{\frac{1}{p_1}} \leq \left(\int_0^T |\overline{u}_{\delta,\Psi}|^{\widetilde{p}}(s)ds \right)^{\frac{\overline{p}}{\widetilde{p}}} T^{\frac{\widetilde{p}-\overline{p}p_1}{\widetilde{p}}}.$$

Arguing as in Corollary 5.13, we know that $\mathcal{F}_\Psi[\overline{u}_{\delta,\Psi}] \leq \mathcal{F}_\Psi[\overline{u}]$ and then

$$\left(\int_0^T |\overline{u}_{\delta,\Psi}|^{\overline{p}p_1}(s)ds \right)^{\frac{1}{p_1}} \leq \left(\int_0^T |\overline{u}|^{\widetilde{p}}(s)ds \right)^{\frac{\overline{p}}{\widetilde{p}}} T^{\frac{\widetilde{p}-\overline{p}p_1}{\widetilde{p}}} =: C_2(T,\widetilde{p},\overline{p},p_1,\overline{u}).$$

Plugging last inequality in Equation (5.9) we get

$$\mathbb{E}\left[\int_{t_1}^{t_2} \frac{e^{-\overline{p}p_1 A(s)}}{G^{\overline{p}p_1}(s)} |\overline{u}_{\delta,\Psi}(s)|^{\overline{p}p_1} ds \right]^{\frac{1}{p_1}} \leq \left(\sup_{t\in[0,T]} e^{\overline{p}(t-A(t))} \right) e^{\frac{\overline{p}(\overline{p}p_1-1)}{2}T} C_2(T,\widetilde{p},\overline{p},p_1,\overline{u}) =: C_3(T,\widetilde{p},\overline{p},p_1,\overline{u}).$$

Setting then $C_4(T,\widetilde{p},\overline{p},p_1,\overline{u}) := C_3(T,\widetilde{p},\overline{p},p_1,\overline{u})C_1(T,\overline{p},p_1)$ we obtain, from Equation (5.8),

(5.10) $$I_2 \leq C_4(T,\widetilde{p},\overline{p},p_1,\overline{u})(t_2-t_1)^{\gamma_1}.$$

Now let us consider I_1. By Hölder's inequality with exponent $\frac{4}{\overline{p}} > 1$ we get

(5.11) $$I_1 \leq \mathbb{E}\left[|G(t_2)e^{A(t_2)} - G(t_1)e^{A(t_1)}|^4 \right]^{\frac{\overline{p}}{4}} \mathbb{E}\left[\left| \int_0^{t_2} \frac{e^{-A(s)}}{G(s)} \overline{u}_{\delta,\Psi}(s)ds \right|^{\frac{4\overline{p}}{4-\overline{p}}} \right]^{\frac{4-\overline{p}}{4}}.$$

Let us first consider the first factor of I_1. Let $Y(t) = G(t)e^{A(t)}$ and observe, by Itô's formula, that

$$dY(t) = a(t)Y(t)dt + Y(t)dW(t)$$

in $[t_1,t_2]$, that is to say

$$Y(t) = Y(t_1) + \int_{t_1}^t a(s)Y(s)ds + \int_{t_1}^t Y(s)dW(s).$$

For any $t \in [0, t_2 - t_1]$ it holds

$$
\begin{aligned}
Y(t + t_1) &= Y(t_1) + \int_{t_1}^{t+t_1} a(s)Y(s)ds + \int_{t_2}^{t+t_2} Y(s)dW(s) \\
&= Y(t_1) + \int_0^t a(s + t_1)Y(s + t_1)ds + \int_0^t Y(s + t_1)d\widetilde{W}(s) \\
&= Y(t_1) + \int_0^t a(s + t_1)(Y(s + t_1) - Y(t_1))ds + \int_0^t a(s + t_1)Y(t_1)ds \\
&\quad + \int_0^t (Y(s + t_1) - Y(t_1))d\widetilde{W}(s) + \int_0^t Y(t_1)d\widetilde{W}(s),
\end{aligned}
$$

where we used the change of variables $s \mapsto s - t_1$ and we set $\widetilde{W}(t) := W(t+t_1) - W(t)$, that is still a Brownian motion. Hence, the process $\widetilde{Y}(t) = Y(t + t_1) - Y(t_1)$ solves the SDE

$$
d\widetilde{Y}(t) = a(t + t_1)\left(\widetilde{Y}(t) + Y(t_1)\right)dt + (\widetilde{Y}(t) + Y(t_1))d\widetilde{W}(t), \quad \widetilde{Y}(0) = 0
$$

with $t \in [0, t_1 - t_2]$. We can extend the process by setting $\widetilde{Y}(t) = Y(t_2) - Y(t_1)$ as $t \in [t_2 - t_1, T]$ so that:

$$
d\widetilde{Y}(t) = \widetilde{a}(t)\left(\widetilde{Y}(t) + Y(t_1)\right)dt + (\widetilde{Y}(t) + Y(t_1))\mathbf{1}_{[0, t_2 - t_1]}d\widetilde{W}(t), \quad \widetilde{Y}(0) = 0,
$$

where $\widetilde{a}(t) = a(t + t_1)$ if $t \in [0, t_2 - t_1]$ and $\widetilde{a}(t) = 0$ if $t \in (t_2 - t_1, T]$. Now set $M = \|a\|_{L^\infty(0,T)} \geq \|\widetilde{a}\|_{L^\infty(0,T)}$ and observe that by Lemma 2.4,

$$
\begin{aligned}
\sup_{t \in [0, t_2 - t_1]} \mathbb{E}[|\widetilde{Y}(t)|^4] &\leq K(2, M, t_2 - t_1)\left(\left(\int_0^{t_2 - t_1} \mathbb{E}[|\widetilde{a}(t)Y(t_1)|^4]^{\frac{1}{4}}dt\right)^4 + \left(\int_0^{t_2 - t_1} \mathbb{E}[|Y(t_1)|^4]^{\frac{1}{2}}dt\right)^2\right) \\
&\leq K(2, M, T)\left(\mathbb{E}[|Y(t_1)|^4]\left(\int_0^{t_2 - t_1} |\widetilde{a}(t)|dt\right)^4 + \mathbb{E}[|Y(t_1)|^4](t_2 - t_1)^2\right) \\
&\leq K(2, M, T)\,\mathbb{E}[|Y(t_1)|^4](t_2 - t_1)^2\left(MT^2 + 1\right),
\end{aligned}
$$

where we used the fact that $T \mapsto K(2, M, T)$ is increasing and $t_2 - t_1 \leq T$. Calling back that

$$
\mathbb{E}[|Y(t_1)|^4] = e^{4A(t_1)}\,\mathbb{E}[G^4(t_1)] \leq \left(\sup_{t \in [0,T]} e^{4A(t)}\right)e^{6T}
$$

by Equation (2.5) and setting $C_5(M, T) := \left(MT^2 + 1\right)K(2, M, T)\left(\sup_{t \in [0,T]} e^{4A(t)}\right)e^{6T}$ we get

$$
\mathbb{E}[|Y(t_2) - Y(t_1)|^4] = \mathbb{E}[|\widetilde{Y}(t_2 - t_1)|^4] \leq \sup_{t \in [0, t_1 - t_2]} \mathbb{E}[|\widetilde{Y}(t)|^4] \leq C_5(M, T)(t_2 - t_1)^2.
$$

Going back to Equation (5.11), setting $\gamma_2 = \frac{\overline{p}}{2}$, where $\gamma_2 > 1$ since $\overline{p} > 2$, we have

$$
\text{(5.12)} \qquad I_1 \leq (C_5(M, T))^{\frac{\overline{p}}{4}}(t_2 - t_1)^{\gamma_2}\,\mathbb{E}\left[\left|\int_0^{t_2} \frac{e^{-A(s)}}{G(s)}\overline{u}_{\delta,\Psi}(s)ds\right|^{\frac{4\overline{p}}{4 - \overline{p}}}\right]^{\frac{4 - \overline{p}}{4}}.
$$

Now we have to estimate the second factor. Let us first use Jensen's inequality (observing that $\frac{4\overline{p}}{4 - \overline{p}} > 1$) to get

$$
\begin{aligned}
\mathbb{E}\left[\left|\int_0^{t_2} \frac{e^{-A(s)}}{G(s)}\overline{u}_{\delta,\Psi}(s)ds\right|^{\frac{4\overline{p}}{4 - \overline{p}}}\right] &\leq t_2^{\frac{4\overline{p}}{4 - \overline{p}} - 1}\,\mathbb{E}\left[\int_0^{t_2} \frac{e^{-\frac{4\overline{p}}{4 - \overline{p}}A(s)}}{G^{\frac{4\overline{p}}{4 - \overline{p}}}(s)}|\overline{u}_{\delta,\Psi}(s)|^{\frac{4\overline{p}}{4 - \overline{p}}}ds\right] \\
&\leq T^{\frac{4\overline{p}}{4 - \overline{p}} - 1}\,\mathbb{E}\left[\int_0^T e^{\frac{4\overline{p}}{4 - \overline{p}}(s - A(s))}|G'(s)|^{\frac{4\overline{p}}{4 - \overline{p}}}|\overline{u}_{\delta,\Psi}(s)|^{\frac{4\overline{p}}{4 - \overline{p}}}ds\right] \\
&\leq T^{\frac{4\overline{p}}{4 - \overline{p}} - 1}\left(\sup_{t \in [0,T]} e^{\frac{4\overline{p}}{4 - \overline{p}}(s - A(s))}\right)\mathbb{E}\left[\sup_{s \in [0,T]} |G'(s)|^{\frac{4\overline{p}}{4 - \overline{p}}}\right]\int_0^T |\overline{u}_{\delta,\Psi}(s)|^{\frac{4\overline{p}}{4 - \overline{p}}}ds.
\end{aligned}
$$

By Doob's maximal inequality we have

$$\mathbb{E}\left[\sup_{s\in[0,T]}|G'(s)|^{\frac{4\overline{p}}{4-\overline{p}}}\right] \leq \left(\frac{4\overline{p}}{5\overline{p}-4}\right)^{\frac{4\overline{p}}{4-\overline{p}}} e^{\frac{2\overline{p}}{4-\overline{p}}\left(\frac{4\overline{p}}{4-\overline{p}}-1\right)T}.$$

On the other hand, let us observe that, being $\overline{p} < \frac{4\widetilde{p}}{4+\widetilde{p}}$, it holds $\frac{\widetilde{p}(4-\overline{p})}{4\overline{p}} > 1$, thus we can use it as exponent for Hölder's inequality, obtaining

$$\int_0^T |\overline{u}_{\delta,\Psi}(s)|^{\frac{4\overline{p}}{4-\overline{p}}}ds \leq \left(\int_0^T |\overline{u}_{\delta,\Psi}(s)|^{\widetilde{p}}ds\right)^{\frac{4\overline{p}}{\widetilde{p}(4-\overline{p})}} T^{1-\frac{4\overline{p}}{\widetilde{p}(4-\overline{p})}} \leq \left(\int_0^T |\overline{u}(s)|^{\widetilde{p}}ds\right)^{\frac{4\overline{p}}{\widetilde{p}(4-\overline{p})}} T^{1-\frac{4\overline{p}}{\widetilde{p}(4-\overline{p})}}.$$

Hence we get, from Equation (5.12),

$$(5.13) \qquad I_1 \leq (C_5(M,T))^{\frac{\overline{p}}{4}}(t_1-t_2)^{\gamma_2}T^{\overline{p}-\frac{\overline{p}}{p}}\left(\sup_{t\in[0,T]} e^{-\overline{p}A(s)}\right)\left(\frac{4\overline{p}}{5\overline{p}-4}\right)^{\overline{p}} e^{\frac{\overline{p}}{2}\left(\frac{4\overline{p}}{4-\overline{p}}+1\right)T}\left(\int_0^T |\overline{u}(s)|^{\widetilde{p}}ds\right)^{\frac{\overline{p}}{\widetilde{p}}}$$

$$=: C_6(M,T,\widetilde{p},\overline{p},p_1,\overline{u})(t_2-t_1)^{\gamma_2}.$$

Now set $\gamma = \min\{\gamma_1,\gamma_2\} > 1$ and combine Equations (5.10) and (5.13) with (5.7) to conclude that there exists $C_7(M,T,\widetilde{p},\overline{p},p_1,\overline{u}) > 0$ (notice that C_7 does not depend on t_1,t_2) such that

$$\mathbb{E}[|X_\delta(t_2)-X_\delta(t_1)|^{\overline{p}}] \leq C_7(M,T,\widetilde{p},\overline{p},p_1,\overline{u})(t_2-t_1)^{\gamma}.$$

This, together with the fact that $X_\delta(0) = 0$ for any $\delta > 0$, guarantees that $\{X_\delta\}_{\delta>0}$ is tight (see, for instance, [43, Theorem 11.6.5]). Thus, by a Corollary of Prohorov's theorem (see, for instance, [43, Corollary 11.6.2]), we know that $X_{\delta_n} \Rightarrow \mathcal{S}_{X_0}\overline{u}$ in $C([0,T])$, concluding the proof. $\qquad\square$

Remark 5.16. The previous theorem clearly holds even if $\Psi(t) = Ct^{\widetilde{p}}$ for some constant $C > 0$.

Theorem 5.15 guarantees that even if $\overline{u}_{\delta,\Psi}$ does not converge strongly to $\overline{u}$ (due, for instance, to a highly oscillatory behaviour), it can be still used to approximate the process $\mathcal{S}_{X_0}\overline{u}$. This comes in handy in the application context, whenever one has to numerically determine some functional properties of $\mathcal{S}_{X_0}\overline{u}$. Moreover, in the proof of the previous theorem, we have also shown that if $\Psi \in \Delta_2$ is a Young function and $\|\overline{u}_{\delta,\Psi}\|_{L^{\widetilde{p}}(0,T)}$ is uniformly bounded for some $\widetilde{p} > 2\sqrt{2}$, then the family $\{X_\delta\}_{\delta>0}$ is tight. This means that, in this case, by Prohorov's theorem, $\{X_\delta\}_{\delta>0}$ is relatively compact, i.e. there exists a process $\overline{X}$ with a.s. continuous sample paths and a sequence $\delta_n \to 0$ such that $X_{\delta_n} \Rightarrow \overline{X}$. Combining the latter observation with Theorem 5.11 we have that not only in this case J admits a minimizer, but there exists a sequence $\delta_n \to 0$ such that $\overline{u}_{\delta_n,\Psi} \rightharpoonup \overline{u}$, $X_{\delta_n} \Rightarrow \overline{X}$ and $\overline{X} = \mathcal{S}_{X_0}\overline{u}$.

Let us now exhibit a necessary and sufficient condition for a function $\overline{u}_{\delta,\Psi}$ to be a minimizer of $J_{\delta,\Psi}$, given in terms of an Euler-Lagrange type equation.

Theorem 5.17. *Let hypotheses (H1) to (H5) hold and Ψ be a Young function with strictly increasing continuous derivative ψ. Then $\overline{u}_{\delta,\Psi}$ is the unique solution of*

$$(5.14) \qquad \delta\frac{\overline{u}_{\delta,\Psi}(t_0)}{|\overline{u}_{\delta,\Psi}(t_0)|}\psi(|\overline{u}_{\delta,\Psi}(t_0)|) = \int_{t_0}^T \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t,\xi_{\overline{u}_{\delta,\Psi}}(t))e^{A(t)-A(t_0)}\frac{G(t)}{G(t_0)}\right]dt, \ \forall t_0 \in [0,T].$$

Proof. The proof follows as in Theorem 3.2. Precisely, let $E_{\overline{u}}$ be the set of Lebesgue points of $\overline{u}_{\delta,\Psi}$ in $(0,T)$, E_L the set of Lebesgue points of L in $(0,T)$ and $E = E_{\overline{u}} \cup E_L$. Let $t_0 \in E$, fix a real number $u \in \mathbb{R}$ and $\varepsilon_0 > 0$ small enough to have $\left(t_0 - \frac{\varepsilon_0}{2}, t_0 + \frac{\varepsilon_0}{2}\right) \subset (0,T)$. Define, for any $\varepsilon \in (0,\varepsilon_0)$, $I_\varepsilon := \left(t_0 - \frac{\varepsilon}{2}, t_0 + \frac{\varepsilon}{2}\right)$,

$$u_\varepsilon(t) = \begin{cases} u & t \in I_\varepsilon \\ \overline{u}_{\delta,\Psi}(t) & \text{otherwise,} \end{cases}$$

$g_1(\varepsilon) := J_{\delta,\Psi}[u_\varepsilon]$, $g_2(\varepsilon) := J[u_\varepsilon]$ and $g_3(\varepsilon) := \mathcal{F}_\Psi[u_\varepsilon]$, so that $g_1 = g_2 + \delta g_3$ and

$$\frac{g_1(\varepsilon)-g_1(0)}{\varepsilon} = \frac{g_2(\varepsilon)-g_2(0)}{\varepsilon} + \delta\frac{g_3(\varepsilon)-g_3(0)}{\varepsilon}.$$

Let us only study the second incremental ratio. It holds

$$\frac{g_3(\varepsilon)-g_3(0)}{\varepsilon} = \frac{1}{\varepsilon}\int_0^T (\Psi(u_\varepsilon(t)) - \Psi(\overline{u}_{\delta,\Psi}(t)))dt = \frac{1}{\varepsilon}\int_{I_\varepsilon} (\Psi(u) - \Psi(\overline{u}_{\delta,\Psi}(t)))dt.$$

Being $\Psi \in \Delta_2$, $\Psi(u) - \Psi(\overline{u}_{\delta,\Psi}(t))$ belongs to $L^1(0,T)$, hence we can consider E_Ψ as the set of Lebesgue points of $\Psi(\overline{u}_{\delta,\Psi}(t))$ and $E' = E \cap E_\Psi$. From now on, let us assume that $t_0 \in E'$. Taking the limit we obtain

$$\lim_{\varepsilon \to 0} \frac{g_3(\varepsilon) - g_3(0)}{\varepsilon} = \Psi(u) - \Psi(\overline{u}_{\delta,\Psi}(t_0)).$$

On the other hand, we have, from Equation (3.13),

$$\lim_{\varepsilon \to 0} \frac{g_2(\varepsilon) - g_2(0)}{\varepsilon} = -(u - \overline{u}_{\delta,\Psi}(t_0)) \int_{t_0}^T \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}_{\delta,\Psi}}(t))e^{A(t)-A(t_0)}\frac{G(t)}{G(t_0)}\right] dt,$$

with the notation introduced in Theorem 3.2. Hence we conclude that

$$\lim_{\varepsilon \to 0} \frac{g_1(\varepsilon) - g_1(0)}{\varepsilon} = \delta\Psi(u) - \delta\Psi(\overline{u}_{\delta,\Psi}(t_0)) - (u - \overline{u}_{\delta,\Psi}(t_0)) \int_{t_0}^T \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}_{\delta,\Psi}}(t))e^{A(t)-A(t_0)}\frac{G(t)}{G(t_0)}\right] dt.$$

However, we know that 0 is a minimum point of g_1, thus it holds

$$(5.15) \qquad \delta\Psi(|u|) - \delta\Psi(|\overline{u}_{\delta,\Psi}(t_0)|) - (u - \overline{u}_{\delta,\Psi}(t_0)) \int_{t_0}^T \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}_{\delta,\Psi}}(t))e^{A(t)-A(t_0)}\frac{G(t)}{G(t_0)}\right] dt \geq 0.$$

Setting

$$H(u) := -\delta\Psi(|u|) + u \int_{t_0}^T \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}_{\delta,\Psi}}(t))e^{A(t)-A(t_0)}\frac{G(t)}{G(t_0)}\right] dt,$$

by Equation 5.15 it holds $\max_{u \in \mathbb{R}} H(u) = H(\overline{u}_{\delta,\Psi}(t_0))$. By Remark 5.5 we know that H is differentiable and then by Fermat's theorem $H'(\overline{u}_{\delta,\Psi}(t_0)) = 0$, that is to say

$$\delta\frac{\overline{u}_{\delta,\Psi}(t_0)}{|\overline{u}_{\delta,\Psi}(t_0)|}\psi(|\overline{u}_{\delta,\Psi}(t_0)|) = \int_{t_0}^T \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}_{\delta,\Psi}}(t))e^{A(t)-A(t_0)}\frac{G(t)}{G(t_0)}\right] dt, \ \forall t_0 \in E'.$$

Finally, the right-hand side being continuous (as we have shown in the proof of Theorem 3.2), we achieve Equation (5.14).

Now we have to show that the latter is also a sufficient condition. To do this, let $\overline{u}_{\delta,\Psi} \in L^1(0,T)$ be a solution of its and $u \in L^1(0,T)$ be any other function. If $u \notin L^\Psi(0,T)$, then $+\infty = J_{\delta,\Psi}[u] \geq J_{\delta,\Psi}[\overline{u}_{\delta,\Psi}]$. Thus, let us consider $u \in L^\Psi(0,T)$. Observe that

$$J_{\delta,\Psi}[u] - J_{\delta,\Psi}[\overline{u}_{\delta,\Psi}] = J[u] - J[\overline{u}_{\delta,\Psi}] + \delta(\mathcal{F}_\Psi[u] - \mathcal{F}_\Psi[\overline{u}_{\delta,\Psi}]).$$

We already know, as it is shown in Theorem 4.1, that

$$(5.16) \qquad J[u] - J[\overline{u}_{\delta,\Psi}] \geq \int_0^T (u(s) - \overline{u}_{\delta,\Psi}(s))\left(-\int_s^T \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}_{\delta,\Psi}}(t))e^{A(t)-A(s)}\frac{G(t)}{G(s)}\right] dt\right) ds.$$

Moreover, by the convexity of Ψ, we get

$$\Psi(|u(s)|) - \Psi(|\overline{u}_{\delta,\Psi}(s)|) \geq \frac{\overline{u}_{\delta,\Psi}(s)}{|\overline{u}_{\delta,\Psi}(s)|}\psi(|\overline{u}_{\delta,\Psi}(s)|)(u(s) - \overline{u}_{\delta,\Psi}(s)), \ \forall s \in (0,T),$$

and then

$$(5.17) \qquad \mathcal{F}_\Psi[u] - \mathcal{F}_\Psi[\overline{u}_{\delta,\Psi}] \geq \int_0^T \frac{\overline{u}_{\delta,\Psi}(s)}{|\overline{u}_{\delta,\Psi}(s)|}\psi(|\overline{u}_{\delta,\Psi}(s)|)(u(s) - \overline{u}_{\delta,\Psi}(s)) ds.$$

Combining Equation (5.16) and (5.17) we get

$$J_{\delta,\Psi}[u] - J_{\delta,\Psi}[\overline{u}_{\delta,\Psi}] \geq \int_0^T (u(s) - \overline{u}_{\delta,\Psi}(s))$$
$$\times \left(\frac{\overline{u}_{\delta,\Psi}(s)}{|\overline{u}_{\delta,\Psi}(s)|}\psi(|\overline{u}_{\delta,\Psi}(s)|) - \int_s^T \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}_{\delta,\Psi}}(t))e^{A(t)-A(s)}\frac{G(t)}{G(s)}\right] dt\right) ds = 0,$$

that is to say that $\overline{u}_{\delta,\Psi}$ is a minimizer for $J_{\delta,\Psi}$ in $L^1(0,T)$.

Hence we conclude that any solution of Equation (5.14) is a minimizer for $J_{\delta,\Psi}$ in $L^1(0,T)$. However, being Ψ strictly convex, $\overline{u}_{\delta,\Psi}$ is the unique minimizer of $J_{\delta,\Psi}$. Thus $\overline{u}_{\delta,\Psi}$ is the unique solution of equation (5.14). □

Remark 5.18. Let us stress out, as we did for J, that $\mathcal{F}_\Psi$ is Gateaux-differentiable in $L^\Psi(0,T)$ with Gateaux derivative

$$\partial_u \mathcal{F}[v] = \int_0^T \frac{u(s)v(s)}{|u(s)|}\psi(|u(s)|)ds, \ v \in L^\Psi(0,T).$$

One can show that the previous quantity is finite for any $v \in L^\Psi(0,T)$ by means of Hölder's inequality for Orlicz spaces [37, Theorem 4.7.5]. As a consequence, we obtain that also $J_{\delta,\Psi}[u]$ is Gateaux-differentiable in $L^\Psi(0,T)$ with Gateaux derivative

$$\partial_u J_{\delta,\Psi}[v] = \partial_u J[v] + \delta\partial_u \mathcal{F}_\Psi[v]$$

and Equation (5.14) follows from Fermat's theorem. For this reason, we can refer to (5.14) as the Euler-Lagrange equation for $J_{\delta,\Psi}$. Let us emphasize that the function $H(u)$ defined in the proof of Theorem 5.17 is, in some sense, the Hamiltonian function associated to $J_{\delta,\Psi}$.

With a suitable choice of the Young function Ψ, we can guarantee better regularity for $\overline{u}_{\delta,\Psi}$.

Corollary 5.19. *Let hypotheses (H1) to (H5) hold and set $\Psi(t) = \frac{t^{2n}}{2n}$ for any $n = 1, 2, 3, \ldots$. Then $\overline{u}_{\delta,\Psi}$ is continuous.*

Proof. With this choice, we have $\psi(t) = t^{2n-1}$ and then Equation (5.14) becomes

$$\delta\frac{\overline{u}_{\delta,\Psi}(t_0)}{|\overline{u}_{\delta,\Psi}(t_0)|}|\overline{u}_{\delta,\Psi}(t_0)|^{2n-1} = \int_{t_0}^T \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}_{\delta,\Psi}}(t))e^{A(t)-A(t_0)}\frac{G(t)}{G(t_0)}\right]dt, \ \forall t_0 \in [0,T],$$

that can be recast as

$$\overline{u}_{\delta,\Psi}(t_0) = \sqrt[2n-1]{\frac{1}{\delta}\int_{t_0}^T \mathbb{E}\left[\frac{\partial F}{\partial \xi}(t, \xi_{\overline{u}_{\delta,\Psi}}(t))e^{A(t)-A(t_0)}\frac{G(t)}{G(t_0)}\right]dt}, \ \forall t_0 \in [0,T].$$

Being the right-hand side continuous, as proved in Theorem 3.2, we conclude the proof. □

6. EXAMPLES

In this section we provide some examples, to highlight on one hand some expected features of the approximation problem while, on the other hand, show some unexpected behaviours even in the easier cases. First we consider the general case of power costs. Then we will focus on the quadratic cost, that is to say the mean squared error approximation functional. In this specific case we are able to restate the Euler-Lagrange equation as a first kind Fredholm equation. The latter property allows us to give some explicit examples via numerical methods.

6.1. **Power cost functionals.** Let us consider $F^{(p)}(\xi) = \frac{|\xi|^p}{p}$ for any $p \geq 2$ and the cost functional $J^{(p)}[u] = \mathbb{E}[\int_0^T F^{(p)}(\xi_u(t))dt]$. For such cost functionals, we are able to prove the following Proposition.

Proposition 6.1. *Fix $p \geq 2$ and let $z \in \mathcal{L}_{\widetilde{p}}^2(\Omega, \mathbb{P}; [0,T])$ for some $\widetilde{p} > p$. Then $J^{(p)}$ satisfies hypotheses (H1) to (H5+).*

Proof. Being $z \in \mathcal{L}_{\widetilde{p}}^2(\Omega, \mathbb{P}; [0,T])$, hypothesis (H1) is satisfied with exponent $\widetilde{p} > p \geq 2$. Clearly, $F^{(p)}(\xi) \geq 0$ and hypothesis (H2) is satisfied. Moreover, since $p \geq 2$, $F^{(p)}$ is twice continuously differentiable in ξ with

$$\frac{dF^{(p)}}{d\xi}(\xi) = |\xi|^{p-2}\xi \qquad\qquad \frac{d^2F^{(p)}}{d\xi^2}(\xi) = (p-1)|\xi|^{p-2},$$

thus we get hypothesis (H3). Next, there exists a constant $L > 0$ such that

$$F^{(p)}(\xi) + \left|\frac{dF^{(p)}}{d\xi}(\xi)\right| + \left|\frac{d^2F^{(p)}}{d\xi^2}(\xi)\right| \leq L(1 + |\xi|^p).$$

As $p < \widetilde{p}$, we achieve hypothesis (H4) with exponent $\alpha = p$. Finally, hypothesis (H5+) is satisfied due to the fact that $F^{(p)}(\xi)$ is strictly convex. □

The Euler-Lagrange equation (3.1) for the functional $J^{(p)}$ can be stated as

$$(6.1) \qquad \int_t^T \mathbb{E}\left[|\xi_{\overline{u}}(\tau)|^{p-2}\xi_{\overline{u}}(\tau)e^{A(\tau)-A(t)}\frac{G(\tau)}{G(t)}\right]d\tau = 0, \ \forall t \in [0,T].$$

To introduce a penalization on the functional $J^{(p)}$, let us consider $\Psi^{(2)}(x) = \frac{x^2}{2}$, $\mathcal{F}^{(2)} := \mathcal{F}_{\Psi^{(2)}}$ and $J_\delta^{(p,2)} := J^{(p)} + \delta\mathcal{F}^{(2)}$. Its Euler-Lagrange equation (5.14) can be recast as

$$(6.2) \qquad \overline{u}_\delta(t) = \frac{1}{\delta}\int_t^T \mathbb{E}\left[|\xi_{\overline{u}_\delta}(\tau)|^{p-2}\xi_{\overline{u}_\delta}(\tau)e^{A(\tau)-A(t)}\frac{G(\tau)}{G(t)}\right]d\tau, \ \forall t \in [0,T].$$

Due to the nature of such equation, we speculate that, for $p > 2$, iteration methods to obtain its solution could be developed. Further investigation on the topic is needed.

In the next subsections, we will focus on the case $p = 2$, in which, as we said before, we are able to restate both equations (6.1) and (6.2) in a more tractable form.

6.2. **The least mean squared error approximation: reduction to Fredholm equations.** Indeed, let us mainly focus on the case $F^{(2)}(\xi) = \frac{\xi^2}{2}$, i.e. the least mean squared error approximation. In particular, let us denote by $J^{(2)}$ the cost functional defined as

$$J^{(2)}[u] = \mathbb{E}\left[\int_0^T \frac{\xi_u^2(t)}{2}dt\right].$$

By Proposition 6.1 we know that, if $z \in \mathcal{L}_p^2(\Omega,\mathbb{P};[0,T])$ for any $p > 2$, then hypotheses $(H1)$ to $(H5+)$ are satisfied and the Euler-Lagrange equation is given by (6.1). Actually, in this case, we can restate the equation as a Fredholm integral equation of the first kind.

Proposition 6.2. Let $z \in \mathcal{L}_p^2(\Omega,\mathbb{P};[0,T])$ for some $p > 2$. Then $\overline{u} \in L^1(0,T)$ is the unique solution of the minimization problem

$$(6.3) \qquad J^{(2)}[\overline{u}] = \min_{u \in L^1(0,T)} J^{(2)}[u]$$

if and only if

$$(6.4) \qquad \int_0^T k(t_0,s;a)\overline{u}(s)ds = \mathcal{Z}(t_0), \ \forall t_0 \in [0,T],$$

where

$$(6.5) \qquad k(t,s;a) = e^{-A(t)-A(s)-\max\{t,s\}}\int_{\max\{t,s\}}^T e^{2A(\tau)+\tau}d\tau$$

and

$$(6.6) \qquad \mathcal{Z}(t) = e^{-A(t)}\int_t^T \int_0^\tau e^{2A(\tau)-A(s)}\mathbb{E}\left[\frac{G^2(\tau)}{G(t)G(s)}z(s)\right]ds\,d\tau.$$

Proof. We already know that $\overline{u}$ is the unique solution of the minimization problem (6.3) if and only if it solves

$$(6.7) \qquad \int_{t_0}^T \mathbb{E}\left[\frac{\xi_{\overline{u}}(t)G(t)}{G(t_0)}e^{(A(t)-A(t_0))}\right]dt = 0, \ \forall t_0 \in [0,T],$$

that is Equation (6.1). By the explicit definition of solution map given in Equation (2.2) we get

$$\xi_{\overline{u}}(t) = \mathcal{S}_0(z-u)(t) = G(t)e^{A(t)}\int_0^t \frac{e^{-A(s)}}{G(s)}(z(s)-\overline{u}(s))ds$$

and then Equation (6.7) becomes

$$(6.8) \qquad \int_{t_0}^T \mathbb{E}\left[\int_0^t \frac{G^2(t)}{G(t_0)G(s)}e^{2A(t)-A(t_0)-A(s)}(z(s)-\overline{u}(s))ds\right]dt = 0, \ \forall t_0 \in [0,T].$$

Now we want to show that we are under the hypotheses of Fubini's theorem, so to exchange the order of the inner integral and the expectation operator. Let us first rewrite

$$\int_0^t \frac{G^2(t)}{G(t_0)G(s)} e^{2A(t)-A(t_0)-A(s)} |z(s) - \overline{u}(s)| ds = \int_0^t \frac{G^2(t)G'(s)}{G(t_0)} e^{2A(t)-A(t_0)-A(s)+s} |z(s) - u(s)| ds$$

$$\leq \left(\sup_{\tau_1,\tau_2,\tau_3 \in [0,T]} e^{2A(\tau_1)-A(\tau_2)-A(\tau_3)+\tau_3} \right) \left(\sup_{s \in [0,T]} G'(s) \right) \frac{G^2(t)}{G(t_0)} \left(\int_0^T |z(s)| ds + \|\overline{u}\|_{L^1(0,T)} \right).$$

Next, we take the expectation on both sides of the previous inequality to achieve

$$\mathbb{E} \left[\int_0^t \frac{G^2(t)}{G(t_0)G(s)} e^{2A(t)-A(t_0)-A(s)} |z(s) - u(s)| ds \right]$$

$$\leq C \mathbb{E} \left[\left(\sup_{s \in [0,T]} G'(s) \right) \frac{G^2(t)}{G(t_0)} \left(\int_0^T |z(s)| ds + \|\overline{u}\|_{L^1(0,T)} \right) \right].$$

By the Cauchy-Schwartz inequality we get

(6.9)
$$\mathbb{E} \left[\left(\sup_{s \in [0,T]} G'(s) \right) \frac{G^2(t)}{G(t_0)} \left(\int_0^T |z(s)| ds + \|u\|_{L^1(0,T)} \right) \right]$$

$$\leq \mathbb{E} \left[\left(\sup_{s \in [0,T]} G'(s) \right)^2 \frac{G^4(t)}{G^2(t_0)} \right]^{\frac{1}{2}} \mathbb{E} \left[\left(\int_0^T |z(s)| ds + \|\overline{u}\|_{L^1(0,T)} \right)^2 \right]^{\frac{1}{2}}.$$

To argue with the first factor of Equation (6.9), let us apply again the Cauchy-Schwartz inequality to obtain

$$\mathbb{E} \left[\left(\sup_{s \in [0,T]} G'(s) \right)^2 \frac{G^4(t)}{G^2(t_0)} \right] \leq \mathbb{E} \left[\left(\sup_{s \in [0,T]} G'(s) \right)^4 \right]^{\frac{1}{2}} \mathbb{E} \left[\frac{G^8(t)}{G^4(t_0)} \right]^{\frac{1}{2}} \leq C \mathbb{E} \left[\frac{G^8(t)}{G^4(t_0)} \right]^{\frac{1}{2}},$$

where we also used Lemma 2.8. Noticing that

$$\frac{G^8(t)}{G^4(t_0)} = e^{8W(t)-4W(t_0)-4t+2t_0},$$

we recall that $\frac{G^8(t)}{G^4(t_0)}$ is a lognormal random variable and then $\mathbb{E} \left[\frac{G^8(t)}{G^4(t_0)} \right]$ is finite. Hence, the first factor of Equation (6.9) is finite.
Concerning the second factor, it clearly holds

$$\mathbb{E} \left[\left(\int_0^T |z(s)| ds + \|\overline{u}\|_{L^1(0,T)} \right)^2 \right] \leq 2 \left(\mathbb{E} \left[\left(\int_0^T |z(s)| ds \right)^2 \right] + \|\overline{u}\|_{L^1(0,T)}^2 \right) < +\infty,$$

since $\mathcal{L}_p^2(\Omega, \mathbb{P}; [0,T]) \subset \mathcal{L}_1^2([0,T]; \Omega, \mathbb{P})$. Hence, we can use Fubini's theorem to rewrite Equation (6.8) as

$$\int_{t_0}^T \int_0^t e^{2A(t)-A(t_0)-A(s)} \mathbb{E} \left[\frac{G^2(t)}{G(t_0)G(s)} (z(s) - \overline{u}(s)) \right] ds dt = 0, \ \forall t_0 \in [0,T],$$

that is equivalent to
(6.10)
$$\int_{t_0}^T \int_0^t e^{2A(t)-A(t_0)-A(s)} \mathbb{E} \left[\frac{G^2(t)}{G(t_0)G(s)} \right] \overline{u}(s) ds dt = \int_{t_0}^T \int_0^t e^{2A(t)-A(t_0)-A(s)} \mathbb{E} \left[\frac{G^2(t)}{G(t_0)G(s)} z(s) \right] ds dt, \ \forall t_0 \in [0,T].$$

Arguing as before, notice that $\frac{G^2(t)}{G(t_0)G(s)}$ is a lognormal random variable with

$$\mathbb{E} \left[\log \left(\frac{G^2(t)}{G(t_0)G(s)} \right) \right] = \mathbb{E} \left[2W(t) - W(t_0) - W(s) - t + \frac{t_0 + s}{2} \right] = \frac{t_0 + s}{2} - t$$

$$\text{Var} \left[\log \left(\frac{G^2(t)}{G(t_0)G(s)} \right) \right] = \mathbb{E} \left[(2W(t) - W(t_0) - W(s))^2 \right] = 4t - 3t_0 - 3s + 2\min\{t_0, s\}$$

where for any random variable $X \in L^2(\Omega, \mathbb{P})$ we set $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$. Hence we get

$$\mathbb{E}\left[\frac{G^2(t)}{G(t_0)G(s)}\right] = e^{t-t_0-s+\min\{t_0,s\}} = e^{t-\max\{t_0,s\}}$$

and then

$$\int_{t_0}^T \int_0^t e^{2A(t)-A(t_0)-A(s)} \mathbb{E}\left[\frac{G^2(t)}{G(t_0)G(s)}\right] \overline{u}(s)dsdt = \int_{t_0}^T \int_0^t e^{2A(t)-A(t_0)-A(s)+t-\max\{t_0,s\}}\overline{u}(s)dsdt.$$

Being $\overline{u} \in L^1(0,T)$, it is clear that we can use Fubini's theorem to achieve

$$\int_{t_0}^T \int_0^t e^{2A(t)-A(t_0)-A(s)+t-\max\{t_0,s\}}\overline{u}(s)dsdt = \int_0^T \left(\int_{\max\{t_0,s\}}^T e^{2A(t)-A(t_0)-A(s)+t-\max\{t_0,s\}}dt\right)\overline{u}(s)ds.$$

Setting $k(t,s;a)$ and $\mathcal{Z}(t)$ as in Equations (6.5) and (6.6) we can rewrite Equation (6.10) as (6.4), concluding the proof. $\qquad\square$

An analogous result can be shown for Equation (6.2).

Proposition 6.3. *Let* $z \in \mathcal{L}_p^2(\Omega, \mathbb{P}; [0,T])$ *for some* $p > 2$. *Then* $\overline{u}_\delta \in L^1(0,T)$ *is the unique solution of the minimization problem*

$$(6.11) \qquad J_\delta^{(2,2)}[\overline{u}_\delta] = \min_{u \in L^1(0,T)} J_\delta^{(2,2)}[u]$$

if and only if

$$(6.12) \qquad \delta\overline{u}_\delta(t_0) + \int_0^T k(t_0,s;a)\overline{u}_\delta(s)ds = \mathcal{Z}(t_0), \ \forall t_0 \in [0,T],$$

where k *and* $\mathcal{Z}$ *are defined in Equations* (6.5) *and* (6.6).

We omit the proof since it is identical to the previous one.
Both Propositions 6.2 and 6.3 give us an alternative form of the Euler-Lagrange equation whose usefulness is twofold: we can use some well-known numerical methods to exploit the solution and it can be also used to determine the existence of the solution (and actually exhibit it) under an additional hypothesis. This is the content of the next subsection.

6.3. **The least mean squared approximation: the independence case.** Now let us prove that if z is independent of W, then we can exhibit the solution of the approximation problem.

Proposition 6.4. *Let* $z \in \mathcal{L}_p^2(\Omega, \mathbb{P}; [0,T])$ *for some* $p > 2$ *be independent of the Brownian motion* $W(t)$. *Then the minimization problem* (6.3) *admits as unique solution* $\overline{u}(t) = \mathbb{E}[z(t)]$. *Moreover, it holds*

$$(6.13) \qquad J^{(2)}[\overline{u}] = \frac{1}{2}\int_0^T \int_0^t \int_0^t e^{2A(t)-A(s)-A(\tau)+t-\max\{s,\tau\}}\mathrm{Cov}(z(s), z(\tau))dsd\tau dt,$$

where Cov *is the covariance operator, i.e., for two random variables* $X, Y \in L^2(\Omega, \mathbb{P})$, $\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$.

Proof. By Proposition 6.2 we know that $\overline{u}$ is the unique solution of the minimization problem (6.3) if and only if it solves equation (6.4). Thus, we only have to show that $\overline{u}(t) = \mathbb{E}[z(t)]$ solves that equation. To do this, just observe that, being z independent of W, by Equation (6.6) we get

$$\mathcal{Z}(t) = e^{-A(t)}\int_t^T \int_0^\tau e^{2A(\tau)-A(s)+\tau-\max\{t,s\}} \mathbb{E}[z(s)]dsd\tau.$$

Since $z \in \mathcal{L}_p^2(\Omega, \mathbb{P}; [0,T])$, we know that $\mathbb{E}[z(\cdot)] \in L^1(0,T)$ and then we can use Fubini's theorem in the previous equation to achieve

$$\mathcal{Z}(t) = \int_0^T \left(\int_{\max t,s}^T e^{-A(t)+2A(\tau)-A(s)+\tau-\max\{t,s\}}d\tau\right) \mathbb{E}[z(s)]ds = \int_0^T k(t,s;a) \mathbb{E}[z(s)]ds,$$

concluding the first part of the proof.

Now let us evaluate $J^{(2)}[\overline{u}]$. We have, by the definition of $J^{(2)}$ and Equation (2.2),

$$
\begin{aligned}
J^{(2)}[\overline{u}] &= \frac{1}{2} \int_0^T \mathbb{E}\left[G^2(t)e^{2A(t)}\left(\int_0^t \frac{e^{-A(s)}}{G(s)}(z(s) - \overline{u}(s))ds\right)^2\right]dt \\
&= \frac{1}{2} \int_0^T \mathbb{E}\left[\int_0^t \int_0^t \frac{G^2(t)e^{2A(t)-A(s)-A(\tau)}}{G(s)G(\tau)}(z(s) - \overline{u}(s))(z(\tau) - \overline{u}(\tau))dsd\tau\right]dt.
\end{aligned}
$$
(6.14)

Now we want to show that we are under the hypotheses of Fubini's theorem so that we can exchange the inner double integral with the expectation operator. To do this, notice that

$$
\int_0^t \int_0^t \frac{G^2(t)e^{2A(t)-A(s)-A(\tau)}}{G(s)G(\tau)}|z(s) - \overline{u}(s)||z(\tau) - \overline{u}(\tau)|dsd\tau
$$

$$
\leq G^2(t)\left(\sup_{t,t_0,s\in[0,T]} e^{2A(t)-A(s)-A(\tau)+s+\tau}\right)\left(\sup_{s\in[0,T]} G'(s)\right)\left(\sup_{s\in[0,T]} G'(\tau)\right)
$$

$$
\times \int_0^t \int_0^t |z(s) - \overline{u}(s)||z(\tau) - \overline{u}(\tau)|dsd\tau.
$$

Taking the expectation on both sides and using the fact that z is independent of W we have

$$
\mathbb{E}\left[\int_0^t \int_0^t \frac{G^2(t)e^{2A(t)-A(s)-A(\tau)}}{G(s)G(\tau)}|z(s) - \overline{u}(s)||z(\tau) - \overline{u}(\tau)|dsd\tau\right]
$$

$$
\leq \mathbb{E}\left[G^2(t)\left(\sup_{t,t_0,s\in[0,T]} e^{2A(t)-A(s)-A(\tau)+s+\tau}\right)\left(\sup_{s\in[0,T]} G'(s)\right)\left(\sup_{s\in[0,T]} G'(\tau)\right)\right.
$$

$$
\times \left.\int_0^t \int_0^t |z(s) - \overline{u}(s)||z(\tau) - \overline{u}(\tau)|dsd\tau\right]
$$

$$
\leq C\,\mathbb{E}\left[G^2(t)\left(\sup_{s\in[0,T]} G'(s)\right)\left(\sup_{s\in[0,T]} G'(\tau)\right)\right]\mathbb{E}\left[\int_0^t \int_0^t |z(s) - \overline{u}(s)||z(\tau) - \overline{u}(\tau)|dsd\tau\right].
$$

Arguing exactly as in the proof of Proposition 6.2, we have

$$
\mathbb{E}\left[G^2(t)\left(\sup_{s\in[0,T]} G'(s)\right)\left(\sup_{s\in[0,T]} G'(\tau)\right)\right] \leq C,
$$

while, on the other hand

$$
\begin{aligned}
\mathbb{E}\left[\int_0^t \int_0^t |z(s) - \overline{u}(s)||z(\tau) - \overline{u}(\tau)|dsd\tau\right] &= \int_0^t \int_0^t \mathbb{E}[|z(s) - \overline{u}(s)||z(\tau) - \overline{u}(\tau)|]dsd\tau \\
&\leq \int_0^t \int_0^t \mathbb{E}[|z(s) - \overline{u}(s)|^2]^{\frac{1}{2}}\,\mathbb{E}[|z(\tau) - \overline{u}(\tau)|^2]^{\frac{1}{2}}dsd\tau \\
&= \int_0^t \mathbb{E}[|z(s) - \overline{u}(s)|^2]^{\frac{1}{2}}ds \int_0^t \mathbb{E}[|z(\tau) - \overline{u}(\tau)|^2]^{\frac{1}{2}}d\tau \\
&\leq \left(\int_0^T \mathbb{E}[|z(s) - \overline{u}(s)|^2]^{\frac{1}{2}}ds\right)^2 \\
&\leq T\int_0^T \mathbb{E}[|z(s) - \overline{u}(s)|^2]ds,
\end{aligned}
$$

where we used the Cauchy-Schwartz inequality and Jensen's inequality. Now let us observe that

$$
\mathbb{E}[|z(s) - \overline{u}(s)|^2] = \mathbb{E}[|z(s)|^2] - \mathbb{E}[z(s)]^2 \leq \mathbb{E}[|z(s)|^2],
$$

and then, being $z \in \mathcal{L}_p^2(\Omega, \mathbb{P}; [0,T])$, we know that

$$
T\int_0^T \mathbb{E}[|z(s) - \overline{u}(s)|^2]ds \leq C.
$$

Hence we can use Fubini's theorem and the fact that z is independent of W in Equation (6.14) to conclude the proof. □

Remark 6.5. Last Proposition agrees, in some sense, with the classical idea that the expected value should minimize the mean squared error under the hypothesis that z is independent of W. However, we cannot remove this hypothesis, as we will see in the next subsection.

Let us use the previous Proposition to provide an example in which we already know that the minimizer exists (and we know its exact form) and we can numerically solve the penalized equations. Precisely, let us set $z(t)$ as a geometric Brownian motion independent of W, so that $\overline{u}(t) = \mathbb{E}[z(t)] \equiv 1$. Let $a(t) \equiv -1$, and then $A(t) = -t$, and $T = 1$. In this case we have

$$k(t, s; -1) = e^{-|t-s|} - e^{-1+\min\{t,s\}} \tag{6.15}$$

and

$$\mathcal{Z}(t) = e^{-1-t}(-e + e^t - 3e^{2t} + 2e^{1+t} + e^{2t}t). \tag{6.16}$$

For any $\delta > 0$, let $\overline{u}_\delta$ be the solution of Equation (6.12). To obtain a numerical evaluation of $\overline{u}_\delta$ for some fixed $\delta > 0$ we used Nyström method (see [6]). Since we expect an highly oscillatory behaviour for small values of δ, we need to determine a big number of nodes. To do this, we adopted a composite Newton-Cotes quadrature formula based on a 7-th order interpolating polynomial on equispaced nodes. Precisely, we divided the interval $[0, 1]$ in N subintervals and, on each interval, we applied the closed Newton-Cotes formula (on equispaced nodes) with weights:

$$w = \frac{1}{840N}(41, 236, 27, 272, 27, 236, 41).$$

To be sure to avoid Runge's phenomenon, one could also reduce the order of the interpolation while increasing N. Such solutions are visualized in Figure 1. Evidently, $\overline{u}_\delta$ does not converge to 1 as $\delta \to 0$. Let us now
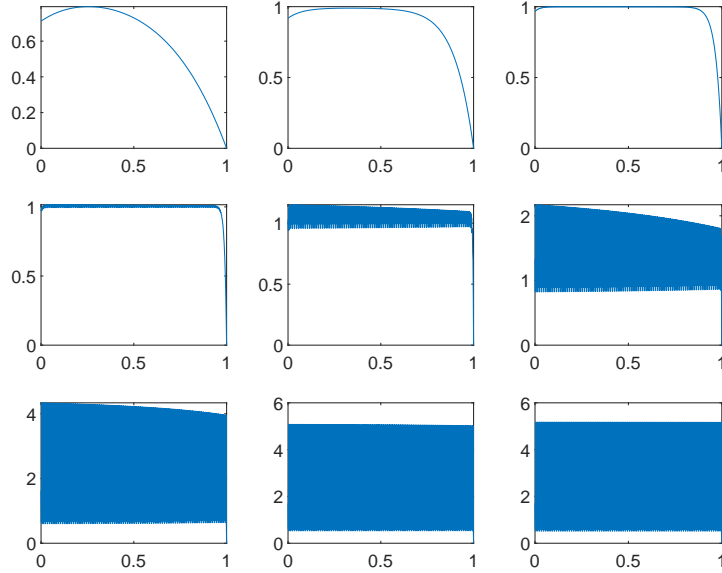

FIGURE 1. Numerical solutions of Equation (6.12) with $T = 1$, $\mathcal{Z}(t)$ given in Equation (6.16) and $k(t, s; a)$ given in Equation (6.15), for different values of δ. Precisely, reading left-to-right top-to-bottom we have $\overline{u}_\delta$ for $\delta = 10^{-n}$ with $n = 1, \ldots, 9$. N is fixed to 100, so that we have 601 nodes for each $\overline{u}_\delta$.

denote $\overline{u}_n = \overline{u}_{10^{-n}}$. To show that $J^{(2)}[\overline{u}_n] \to J^{(2)}[\overline{u}]$, let us first evaluate $J^{(2)}[\overline{u}]$. This can be done by

observing that

$$\text{Cov}(z(s), z(\tau)) = e^{\min\{\tau, s\}} - 1$$

and then, by Equation (6.13),

$$J^{(2)}[\overline{u}] = \frac{1}{2} \int_0^1 \int_0^t \int_0^t e^{-t+\min\{\tau,s\}} (e^{\min\{\tau,s\}} - 1) ds d\tau dt = \frac{e^2 - 7}{4e} \approx 0.0357814.$$

On the other hand, to evaluate $J^{(2)}[\overline{u}_n]$, we adopted a numerical method based on a Monte-Carlo approach. Precisely we simulated a skeleton of $6N + 1$ nodes $(\xi_i)_{0 \le i \le 6N}$ for the process $\xi := \xi_{\overline{u}_n}$. To do this, first we simulated a skeleton $(z_i)_{0 \le i \le 6N}$ of $6N + 1$ nodes for z as

$$\begin{cases} z_0 = 1, \\ z_i = z_{i-1} e^{\frac{\widetilde{\zeta}_i}{\sqrt{6N}} - \frac{1}{12N}}, & i = 1, \dots, 6N, \end{cases}$$

where $\widetilde{\zeta}_i \sim \mathcal{N}(0, 1)$ with $\widetilde{\zeta}_i$ independent of $\widetilde{\zeta}_j$ for each $i \ne j$. Once this is done, $(\xi_i)_{0 \le i \le 6N}$ can be obtained by using an Euler scheme (see [5]):

$$\begin{cases} \xi_0 = 0 \\ \xi_i = \xi_{i-1} + \frac{-\xi_{i-1} + z_{i-1} - \overline{u}_n(t_{i-1})}{6N} + \frac{\zeta_i \xi_{i-1}}{\sqrt{6N}}, & i = 1, \dots, 6N, \end{cases}$$

where $t_i = \frac{i}{6N}$, $\overline{u}_n(t_{i-1})$ has been obtained previously via Nyström method and $\zeta_i \sim \mathcal{N}(0, 1)$ with ζ_i independent of ζ_j for $i \ne j$. The value $\frac{1}{2} \int_0^T \xi^2(t) dt$ is then approximated by a quadrature formula and $J^{(2)}[\overline{u}_n]$ by repeating the procedure for a fixed number N_{traj} of trajectories and then taking the average. While, on one hand, the convergence $J^{(2)}[\overline{u}_n] \to J^{(2)}[\overline{u}]$ is justified by Theorem 5.10, on the other hand the stochastic differential equation could be *stiff* due to the highly oscillatory behaviour of $\overline{u}_n$ and the Euler scheme could fail to catch ξ. The estimated values of $J^{(2)}[\overline{u}_n]$ for $n = 1, \dots, 5$ are given in Table 1.

$J^{(2)}[\overline{u}_1]$	$J^{(2)}[\overline{u}_2]$	$J^{(2)}[\overline{u}_3]$	$J^{(2)}[\overline{u}_4]$	$J^{(2)}[\overline{u}_5]$	$J^{(2)}[\overline{u}]$
0.0462	0.0363	0.0361	0.0362	0.0358	0.0358

TABLE 1. Numerically estimated values of $J^{(2)}[\overline{u}_n]$ for $n = 1, \dots, 5$, in comparison with $J^{(2)}[\overline{u}]$. N is fixed to 100, while $N_{\text{traj}} = 100000$.

To show a numerical evidence that $\overline{u}_n \rightharpoonup \overline{u}$, we also numerically evaluated $\int_0^1 t^j \overline{u}_n(t) dt$ for different values of j and n and we compared it with $\int_0^1 t^j dt = 1/j$ in Table 2.

	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$\int_0^1 x^j dx$
$j = 0$	0.6141	0.8878	0.9670	0.9899	0.9968	0.9988	0.9991	0.9992	0.9992	1
$j = 1$	0.2535	0.4067	0.4690	0.4901	0.4968	0.4988	0.4991	0.4992	0.4992	1/2
$j = 2$	0.1431	0.2495	0.3034	0.3235	0.3301	0.3321	0.3325	0.3325	0.3325	1/3
$j = 3$	0.0930	0.1734	0.2211	0.2403	0.2468	0.2488	0.2491	0.2492	0.2492	1/4

TABLE 2. Numerically estimated values of $\int_0^1 t^j \overline{u}_n(t) dt$ for $n = 1, \dots, 9$ and $j = 0, 1, 2, 3$, in comparison with $\int_0^1 t^j dt = 1/j$. N is fixed to 100 and the values of the integrals are obtained by using the same quadrature formula as applied before to determine $\overline{u}_n$.

With this example, we want to highlight the fact that even if the solution of the minimizing problem (6.3) is known and quite regular, the solution of the penalized problem converge towards them only weakly. However, this is a problem only in the case one wants to approximate the actual minimizer $\overline{u}$. Indeed, usually one is interested in properties of the approximating process $\mathcal{S}_{X_0} \overline{u}$, that, despite the weak convergence of $\overline{u}_\delta$ towards $\overline{u}$, is in the overall approximated well enough by $\mathcal{S}_{X_0} \overline{u}_\delta$, as shown in Theorem 5.15.

As already stated in Remark 6.5, Proposition 6.4 seems to suggest that the expected value should be, in some sense, the minimizer of the mean squared error. However, as we will see in the following example, this is not necessarily true if we suppose that z and W are dependent.

6.4. **The least mean squared error approximation: a dependence case.** Now let us consider a different example. Let $z = G$, $a \equiv 1$, so that $A(t) = t$, and $T = 1$. First of all, let us observe that, since $z(t) = f(t, W(t))$ for some function f, we cannot use Proposition 6.4. Thus, let us first determine (at least numerically) the solutions $\overline{u}_\delta$ of the penalized problem (6.11). According to Equation (6.5) we have

$$(6.17) \qquad k(t, s; 1) = \frac{1}{3} \left(e^{3 - \min\{t,s\} - 2\max\{t,s\}} - e^{|t-s|} \right).$$

Concerning $\mathcal{Z}(t)$, starting from Equation (6.6), it holds

$$(6.18) \qquad \mathcal{Z}(t) = \int_t^1 \int_0^\tau e^{2\tau - t - s} \, \mathbb{E}\left[\frac{G^2(\tau)}{G(t)} \right] ds d\tau.$$

In particular, $\frac{G^2(\tau)}{G(t)} = e^{2W(\tau) - W(t) - \tau + \frac{t}{2}}$ is a lognormal random variable with

$$\mathbb{E}\left[\log\left(\frac{G^2(\tau)}{G(t)} \right) \right] = \mathbb{E}\left[2W(\tau) - W(t) - \tau + \frac{t}{2} \right] = \frac{t}{2} - \tau$$

$$\mathrm{Var}\left[\log\left(\frac{G^2(\tau)}{G(t)} \right) \right] = \mathbb{E}[(2W(\tau) - W(t))^2] = 4\tau - 3t$$

hence

$$\mathbb{E}\left[\frac{G^2(\tau)}{G(t)} \right] = e^{\tau - t}.$$

Thus Equation (6.18) becomes

$$(6.19) \qquad \mathcal{Z}(t) = \int_t^1 \int_0^\tau e^{3\tau - 2t - s} ds d\tau = \frac{1}{6}(3 - 2e^t + e^{2-2t}(2e - 3)).$$

As before, let us exploit some numerical solutions $\overline{u}_\delta$ of Equation (6.12) (with k and $\mathcal{Z}$ given in Equations (6.17) and (6.19)) by using Nyström's method, as shown in Figure 2. From now on let us denote $\overline{u}_n := \overline{u}_{10^{-n}}$. We do not know if Equation (6.4) admits a solution. To have a qualitative idea on whether a solution of
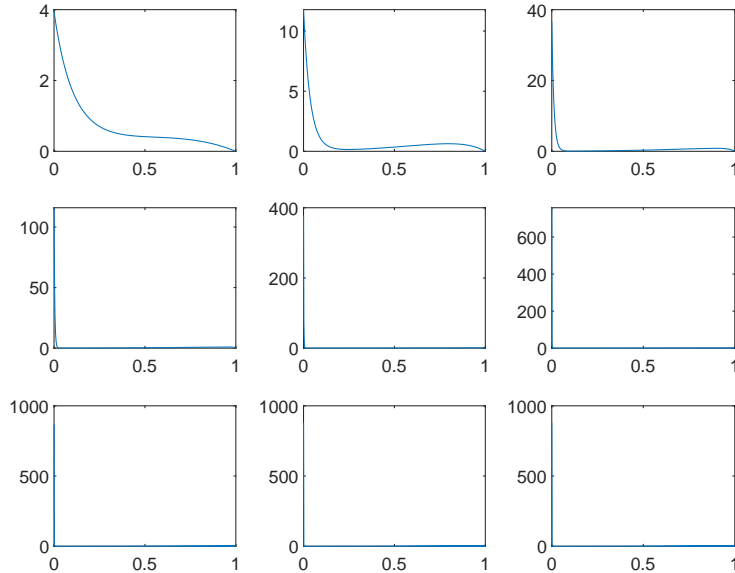

FIGURE 2. Numerical solutions of Equation (6.12) with $T = 1$, $\mathcal{Z}(t)$ given in Equation (6.19) and $k(t, s; a)$ given in Equation (6.17), for different values of δ. Precisely, reading left-to-right top-to-bottom we have $\overline{u}_\delta$ for $\delta = 10^{-n}$ with $n = 1, \dots, 9$. N is fixed to 100, so that we have 601 nodes for each $\overline{u}_\delta$.

	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$
$p=2$	1.0492	2.7499	8.1839	24.5822	86.3839	282.1855	367.4504	378.6386	379.7967
$p=1.5$	0.7900	1.2318	2.0335	3.3589	5.8571	10.7897	12.8737	13.1871	13.2218
$p=1.25$	0.7262	0.9220	1.1764	1.4681	1.8501	2.4278	2.6948	2.7466	2.7529
$p=1.1$	0.7057	0.8139	0.9178	1.0014	1.0824	1.1826	1.2347	1.2472	1.2489
$p=1.01$	0.6996	0.7701	0.8183	0.8402	0.8508	0.8584	0.8622	0.8632	0.8633

TABLE 3. Numerically estimated values of $\|\overline{u}_n(t)\|^p_{L^p(0,1)}$ for $n = 1, \ldots, 9$ and $p = 2, 1.5, 1.25, 1.1, 1.01$. N is fixed to 100 and the values of the integrals are obtained by using the same quadrature formula as before to determine $\overline{u}_n$.

Equation (6.4) exists or not, we could evaluate $\|\overline{u}_n\|_{L^p(0,T)}$ for some $p > 1$, as done in Table 3. From Table 3, we expect that $\|\overline{u}_n\|_{L^p(0,1)} \leq C$ for some suitable choice of $p > 1$ (a good choice could be $p = 1.01$, but, for a big value of C, also $p = 2$ seems to work). This numerical evidence lets us conjecture that $\overline{u}_n$ is uniformly bounded in $L^p(0,1)$ for some $p > 1$ and then, by Theorem 5.11, that a solution $\overline{u}$ of the minimization problem (6.3) exists. Since we can suppose $\|\overline{u}_n\|_{L^2(0,1)} \leq C$, let us conjecture that $\overline{u} \in L^2(0,1)$.

With this idea in mind, let us evaluate numerically the solution of Equation (6.4). To do this, we cannot use Nyström's method, as it is well known that for Fredholm integral equations of the first kind the matrix obtained with the quadrature formula is very ill-conditioned. Hence, we have to use a different method. Precisely, we use a Galërkin-type method as follows (see [44, Section 6.3]). Let $P_n(t)$ be the n-th degree Legendre polynomial and define $Q_n(t) = P_n(2t - 1)$. Thus $\{Q_n(t)\}_{n \in \mathbb{N}}$ constitute an orthogonal system in $L^2(0,1)$ equipped with the usual scalar product $\langle \cdot, \cdot \rangle$, i.e.

$$\langle f, g \rangle = \int_0^1 f(t)g(t)dt, \ \forall f, g \in L^2(0,1).$$

Thus, $\overline{u} = \sum_{i=0}^{+\infty} \langle \overline{u}, Q_i \rangle Q_i$ and $\mathcal{Z} = \sum_{i=0}^{+\infty} \langle \mathcal{Z}, Q_i \rangle Q_i$. Moreover, for fixed $s \in [0,1]$, it holds $k(\cdot, s; 1) = \sum_{i=0}^{+\infty} \langle k(\cdot, s; 1), Q_i \rangle Q_i$. Rewriting Equation (6.4) by using the series decomposition of $\overline{u}$ we have

$$\sum_{i=0}^{+\infty} \langle \overline{u}, Q_i \rangle \int_0^1 k(t, s; 1)Q_i(s)ds = \mathcal{Z}(t).$$

Then, using the decomposition of $\mathcal{Z}$ and $k(\cdot, s; 1)$, we get

$$\left(\sum_{i,j=0}^{+\infty} \langle \overline{u}, Q_i \rangle \int_0^1 \langle k(\cdot, s; 1), Q_j \rangle Q_i(s)ds \right) Q_j(t) = \sum_{j=0}^{+\infty} \langle \mathcal{Z}, Q_j \rangle Q_j(t),$$

that is to say

$$\left(\sum_{i,j=0}^{+\infty} \langle \overline{u}, Q_i \rangle \int_0^1 \int_0^1 k(\tau, s; 1)Q_j(\tau)Q_i(s)d\tau ds \right) Q_j(t) = \sum_{j=1}^{+\infty} \langle \mathcal{Z}, Q_j \rangle Q_j(t).$$

Let us fix $m \in \mathbb{N}$ and consider $\mathcal{Z}_m = \sum_{i=0}^{m} \langle \mathcal{Z}, Q_i \rangle Q_i$, that is a finite-dimensional approximation of $\mathcal{Z}$. Let also $\overline{u}^{(m)} = \sum_{i=0}^{m} \langle \overline{u}, Q_i \rangle Q_i$ be a finite-dimensional approximation of $\overline{u}$ satisfying equation (6.4) with $\mathcal{Z}_m$ in place of $\mathcal{Z}$. Then, if we reduce the problem to finding the finite-dimensional approximation $\overline{u}^{(m)}$, it is equivalent to the problem of solving the following linear system

$$\mathbf{Ku} = \mathbf{z},$$

where $\mathbf{K}_{i,j} = \int_0^1 \int_0^1 k(\tau, s; 1)Q_j(\tau)Q_i(s)d\tau ds$, $\mathbf{u}_i = \langle \overline{u}, Q_i \rangle$ and $\mathbf{z}_i = \langle \mathcal{Z}, Q_i \rangle$, for $i, j = 0, \ldots, m$. The approximation $\overline{u}^{(m)}$ is shown in Figure 3. From now on we will refer to $\overline{u}^{(m)}$ directly as $\overline{u}$. Just looking at the figure, it seems that $\overline{u} \not\equiv 1 \equiv \mathbb{E}[z(\cdot)]$. Indeed, in this case, $\mathbb{E}[z(\cdot)]$ is not a solution of the minimization problem (6.3), since

$$\int_0^1 k(t, s; 1)ds = -\frac{1}{6}e^{-3t}(e^3 - 4e^{3t} + 2e^{4t} - 2e^{3+t} + 3e^{1+2t}) \not\equiv \mathcal{Z}(t),$$
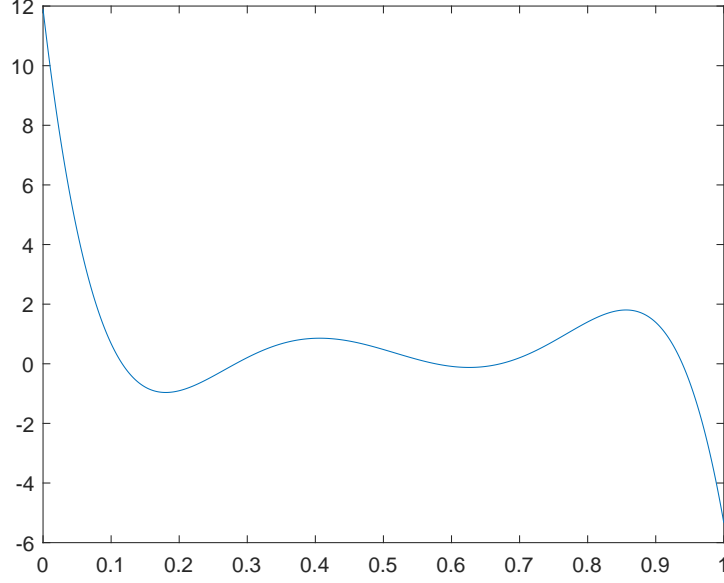
FIGURE 3. The approximation $\overline{u}^{(m)}$, with $m = 5$, of the solution of Equation (6.4) with $k(t, s; a)$ given in Equation (6.17) and $\mathcal{Z}$ given in Equation (6.19).

so that $\mathbb{E}[z(\cdot)]$ does not solve Equation (6.4). Actually, the expected value seems to be quite far from the optimal approximation. This can be observed by evaluating $J^{(2)}[\overline{u}_n]$, $J^{(2)}[\overline{u}]$ and $J^{(2)}[\mathbb{E}[z(\cdot)]]$. As before, to evaluate $J^{(2)}[\overline{u}_n]$ we use the Monte-Carlo approach presented in the previous subsection, with the following set of nodes:

$$\begin{cases} z_0 = 1, \\ z_i = z_{i-1} e^{\frac{\zeta_i}{\sqrt{6N}} - \frac{1}{12N}}, \quad i = 1, \ldots, 6N, \end{cases}$$

and

$$\begin{cases} \xi_0 = 0 \\ \xi_i = \xi_{i-1} + \dfrac{\xi_{i-1} + z_{i-1} - \overline{u}_n(t_{i-1})}{6N} + \dfrac{\zeta_i \xi_{i-1}}{\sqrt{6N}}, \quad i = 1, \ldots, 6N, \end{cases}$$

where $t_i = \frac{i}{6N}$, $\overline{u}_n(t_{i-1})$ has been obtained previously via Nyström's method and $\zeta_i \sim \mathcal{N}(0, 1)$ with ζ_i independent of ζ_j for $i \neq j$. To evaluate $J^{(2)}[\overline{u}]$ and $J^{(2)}[\mathbb{E}[z(\cdot)]]$ we can use the same exact scheme substituting respectively $\overline{u}$ (obtained by Galërkin's method) and 1 in place of $\overline{u}_n$. Again, the evaluations of $J^{(2)}[\overline{u}_n]$ for big values of n are not reliable due to the stiffness of the underlying problem. The results are exposed in Table 4: here it is evident that the expected value is not the optimal approximation.

$J^{(2)}[\overline{u}_1]$	$J^{(2)}[\overline{u}_2]$	$J^{(2)}[\overline{u}_3]$	$J^{(2)}[\overline{u}]$	$J^{(2)}[\mathbb{E}[z(\cdot)]]$
0.0836	0.0572	0.0510	0.0577	0.1505

TABLE 4. Numerically estimated values of $J^{(2)}[\overline{u}_n]$ for $n = 1, \ldots, 3$, in comparison with $J^{(2)}[\overline{u}]$ and $J^{(2)}[\mathbb{E}[z(\cdot)]]$. N is fixed to 100, while $N_{\text{traj}} = 100000$. Consider that, since we are using a Monte-Carlo method, all the values in the table are subject to fluctuations, hence results that are *near* to the best error $J^{(2)}[\overline{u}]$ are still admissible, despite being inferior to it. In some sense, this phenomenon, that is expected due to the stochastic approach used, also evidence the speed of convergence of $J^{(2)}[\overline{u}_n]$ to the best error.

Remark 6.6. In the case of the additive noise, in [4] it has been shown that the expected value is always the optimal approximation with respect to the quadratic cost (even if z and W are not independent). Clearly, the presence of the multiplicative noise has a crucial effect in this sense.

Again, to have another numerical evidence of the fact that $\overline{u}_n \rightharpoonup \overline{u}$, we compare $\int_0^1 t^j \overline{u}_n(t)dt$ with $\int_0^1 t^j \overline{u}(t)dt$ for $j = 0, 1, 2, 3$ in Table 5.

	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$\int_0^1 t^j\overline{u}(t)dt$
$j=0$	0.6992	0.7661	0.8093	0.8262	0.8318	0.8333	0.8336	0.8336	0.8336	0.8336
$j=1$	0.1838	0.2256	0.2657	0.2822	0.2877	0.2893	0.2896	0.2896	0.2896	0.2896
$j=2$	0.0942	0.1510	0.1981	0.2175	0.2240	0.2260	0.2263	0.2264	0.2264	0.2264
$j=3$	0.0606	0.1139	0.1580	0.1768	0.1833	0.1852	0.1856	0.1856	0.1856	0.1856

TABLE 5. Numerically estimated values of $\int_0^1 t^j \overline{u}_n(t)dt$ for $n = 1, \ldots, 9$ and $j = 0, 1, 2, 3$, in comparison with $\int_0^1 t^j \overline{u}(t)dt$. N is fixed to 100 and the values of the integrals are obtained by using the same quadrature formula as applied before to determine $\overline{u}_n$.

Remark 6.7. Let us emphasize that one must pay attention to the choice of the numerical method to solve Equation (6.12). Indeed, one cannot exclude *a priori* an highly oscillatory behaviour of the solution of (6.12), as show by our first example. Thus, if a Galërkin-type method is adopted, then the family of independent functions on $[0, T]$ should be chosen according to the expected behaviour of the solutions.

APPENDIX A. LEBESGUE POINTS

Let us recall the definition of Lebesgue point for a function $f \in L^1(0, T)$.

Definition A.1. *We say that $t \in (0, T)$ is a Lebesgue point for f if*

$$\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \int_{t-\frac{\varepsilon}{2}}^{t+\frac{\varepsilon}{2}} f(\tau)d\tau = f(t).$$

We denote by E_f the set of Lebesgue points of f.

By Lebesgue's differentiation theorem (see [17, Section 1.7, Theorem 1]) it is well known that $|[0, T] \setminus E_f| = 0$. Let us recall, in particular, the following convergence result (see [17, Section 1.7, Corollary 2]).

Proposition A.1. *Let $f \in L^p(0, T)$ for some $1 \le p < \infty$ and t be a Lebesgue point for f. Let $\mathcal{I}(t)$ be the family of all closed intervals in $[0, T]$ containing t. Then*

$$\lim_{\substack{\text{diam}(I) \to 0 \\ I \in \mathcal{I}(t)}} \frac{1}{|I|} \int_I |f(\tau) - f(t)|^p d\tau = 0.$$

We can use last statement to prove the following result.

Proposition A.2. *Consider $1 \le p \le \infty$ and let $f \in L^p(0, T)$ and $g \in L^q(0, T)$ where $\frac{1}{p} + \frac{1}{q} = 1$ and define $h = fg \in L^1(0, T)$. Consider versions of f and g that are everywhere finite. Then $E_f \cap E_g \subseteq E_h$. Moreover, if $p = 1$ and $g \in C([0, T])$, then $E_f \subseteq E_h$.*

Proof. Consider $t \in E_f \cap E_g$ and observe that, by Hölder's inequality, it holds, for any closed interval I containing t,

$$\left| \frac{1}{|I|} \int_I h(\tau)d\tau - h(t) \right| \leq \frac{1}{|I|} \int_I |h(\tau) - h(t)|d\tau$$

(A.1)
$$\leq \frac{1}{|I|} \int_I |f(\tau)||g(\tau) - g(t)|d\tau + \frac{1}{|I|} \int_T |g(t)||f(\tau) - f(t)|d\tau$$

$$\leq \left(\frac{1}{|I|} \int_I |f(\tau)|^p d\tau \right)^{\frac{1}{p}} \left(\frac{1}{|I|} \int_I |g(\tau) - g(t)|^q d\tau \right)^{\frac{1}{q}}$$

$$+ |g(t)| \left(\frac{1}{|I|} \int_I |f(\tau) - f(t)|^p d\tau \right)^{\frac{1}{p}}.$$

Being $|g(t)| < \infty$ and $|f(t)| < \infty$, taking the limit as $\mathrm{diam}(I) \to 0$ in Equation (A.1), it holds

$$\lim_{\substack{\mathrm{diam(I)} \to 0 \\ I \in \mathcal{I}(t)}} \left| \frac{1}{|I|} \int_I h(\tau)d\tau - h(t) \right| = 0,$$

so that $t \in E_h$.

Concerning the second part of the statement, just observe that if g is continuous, $E_g = [0, T]$ by the integral mean value theorem. $\qquad \square$

APPENDIX B. LOWER SEMICONTINUITY OF THE FUNCTIONAL $\mathcal{F}$: PROOF OF PROPOSITION 5.7

Proof. First, let us show that $\mathcal{F}_\Psi$ is lower semicontinuous in any $y \in L^1(0, T)$. Hence, let us consider $y_n \to y$ in L^1: we want to show that

$$\liminf_{n \to +\infty} \mathcal{F}_\Psi[y_n] \geq \mathcal{F}_\Psi[y].$$

Without loss of generality, we can consider a non-relabelled subsequence y_n that realizes the limit inferior. Let us first consider the case $\mathcal{F}_\Psi[y] < +\infty$. Let y_{n_k} be a subsequence of y_n that converges almost everywhere to y. By Egorov's theorem (see [17, Theorem 1.2.3]) we know that for any $\delta > 0$ there exists a compact set H such that $y_{n_k} \to y$ uniformly on H and $|[0, T] \setminus H| < \delta$. Moreover, let us define the measure μ on $[0, T]$ such that for any Lebesgue-measurable set $A \subseteq [0, T]$ it holds

$$\mu(A) = \int_A \Psi(y(t))dt,$$

that is to say the measure μ is defined via $\frac{d\mu}{dt} = \Psi(y(t))$. In particular, μ is absolutely continuous with respect to the Lebesgue measure. Fix $\varepsilon > 0$. By absolute continuity there exists $\delta > 0$ such that for any measurable set $A \subseteq [0, T]$, $|A| < \delta$ implies $\mu(A) < \varepsilon$. Let us consider the compact set H obtained by Egorov's theorem such that $|[0, T] \setminus H| < \delta$. In particular, we get

$$\mathcal{F}_\Psi[y] = \mu(H) + \mu([0, T] \setminus H) < \mu(H) + \varepsilon.$$

Being $\mathcal{F}_\Psi[y] < +\infty$ we have that

$$\mu(H) > \mathcal{F}_\Psi[y] - \varepsilon.$$

On the other hand, it also holds

$$\mathcal{F}_\Psi[y_{n_k}] = \int_0^T \Psi(y_n(t))dt \geq \int_H \Psi(y_{n_k}(t))dt.$$

Since $y_{n_k} \to y$ uniformly on H and Ψ is continuous, we also have $\Psi(y_{n_k}) \to \Psi(y_n)$ uniformly on H and, taking the limit, we achieve

$$\liminf_{n \to +\infty} \mathcal{F}_\Psi[y_n] = \lim_{k \to +\infty} \mathcal{F}_\Psi[y_{n_k}] \geq \int_H \Psi(y(t))dt = \mu(H) > \mathcal{F}_\Psi[y] - \varepsilon.$$

Being $\varepsilon > 0$ arbitrary, we conclude the proof in the case $\mathcal{F}_\Psi[y] < +\infty$.

If $\mathcal{F}_\Psi[y] = +\infty$, let us consider the sequence of measurable sets $S_m = \{t \in [0, T] : \Psi(y(t)) \leq m\}$ for $m \in \mathbb{N}$, so that

$$\int_{S_m} \Psi(y(t))dt \leq mT < +\infty.$$

However, by monotone convergence theorem

$$\lim_{m \to +\infty} \int_{S_m} \Psi(y(t))dt = \mathcal{F}_\Psi[y] = +\infty.$$

Thus, for any $M > 0$ there exists $m \in \mathbb{N}$ such that

$$M < \int_{S_m} \Psi(y(t))dt < +\infty.$$

For any Lebesgue-measurable set $A \subset [0, T]$ define $\sigma_m(A) = |A \cap S_m|$ and

$$\mu_m(A) = \int_{A \cap S_m} \Psi(y(t))dt$$

that are two positive measures with $\mu_m \ll \sigma_m$ and $\sigma_m([0, T]) \leq T < +\infty$. We can argue as before, applying Egorov's theorem to σ_m, to achieve

$$\liminf_{n \to +\infty} \int_{S_m} \Psi(y_n(t))dt \geq \int_{S_m} \Psi(y(t))dt - \varepsilon > M - \varepsilon$$

for any $\varepsilon > 0$. On the other hand

$$\mathcal{F}_\Psi[y_n] \geq \int_{S_m} \Psi(y_n(t))dt$$

and then

$$\liminf_{n \to +\infty} \mathcal{F}_\Psi[y_n] > M - \varepsilon$$

for any $\varepsilon > 0$. Being $\varepsilon > 0$ arbitrary, we get

$$\liminf_{n \to +\infty} \mathcal{F}_\Psi[y_n] \geq M.$$

that leads to

$$\liminf_{n \to +\infty} \mathcal{F}_\Psi[y_n] = +\infty = \mathcal{F}[y].$$

Finally, it is well-known that convex lower semicontinuous functions are also weakly lower semicontinuous, as a consequence of Mazur's theorem (see [15]).   □

References

[1] D. Andersson and B. Djehiche. A maximum principle for SDEs of mean-field type. *Applied Mathematics & Optimization*, 63:341–356, 2011.

[2] F. Angrisani, G. Ascione, C. Leone, and C. Mantegazza. *Appunti di Calcolo delle Variazioni*. Amazon, 2019. Lecture notes, Dipartimento di Matematica e Applicazioni "Renato Caccioppoli" dell'Università Federico II di Napoli.

[3] L. Arnold. *Random Dynamical Systems*. Monographs in Mathematics. Springer, 1998.

[4] G. Ascione, G. D'Onofrio, L. Kostal, and E. Pirozzi. An optimal Gauss–Markov approximation for a process with stochastic drift and applications. *Stochastic Processes and their Applications*, 130(11):6481–6514, 2020.

[5] S. Asmussen and P. W. Glynn. *Stochastic simulation: algorithms and analysis*, volume 57. Springer Science & Business Media, 2007.

[6] K. E. Atkinson. *The Numerical Solution of Integral Equations of the Second Kind*, volume 4. Cambridge University Press, 1997.

[7] E. Bayraktar, A. Cosso, and H. Pham. Randomized dynamic programming principle and Feynman-Kac representation for optimal control of McKean-Vlasov dynamics. *Transactions of the American Mathematical Society*, 370(3):2115–2160, 2018.

[8] R. Bellman. Dynamic programming and stochastic control processes. *Information and Control*, 1(3):228–239, 1958.

[9] M. S. Berger. *Nonlinearity and functional analysis: lectures on nonlinear problems in mathematical analysis*, volume 74. Academic press, 1977.

[10] R. Bhattacharya and M. Majumdar. *Random Dynamical Systems: Theory and Applications*. Cambridge University Press, 2007.

[11] J. F. Bonnans and F. J. Silva. First and second order necessary conditions for stochastic optimal control problems. *Applied Mathematics & Optimization*, 65:403–439, 2012.

[12] A. Braides. *Gamma-convergence for Beginners*, volume 22. Clarendon Press, 2002.

[13] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media, 2010.

[14] F. Cordoni and L. Di Persio. A maximum principle for a stochastic control problem with multiple random terminal times. *Mathematics in Engineering*, 2:557, 2020.

[15] B. Dacorogna. *Direct methods in the calculus of variations*, volume 78. Springer Science & Business Media, 2007.

[16] E. De Giorgi. *Semicontinuity theorems in the calculus of variations*. Quaderni dell'Accademia Pontaniana. Accademia Pontaniana, 2008.

[17] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions*. CRC press, 2015.

[18] O. Faugeras, E. Soret, and E. Tanré. Asymptotic behaviour of a network of neurons with random linear interactions. *arXiv: Probability*, 2019.

[19] F. Flandoli and E. Tonello. An introduction to random dynamical systems for climate. `https://courseclimath19.sciencesconf.org/data/pages/RDS_4_Climate_Flandoli_v2.pdf`, 2019. Online, last accessed on 07/10/2021.

[20] M. Fuhrman, Y. Hu, and G. Tessitore. Stochastic maximum principle for optimal control of SPDEs. *ArXiv*, abs/1302.0286, 2013.

[21] M. Fuhrman and C. Orrieri. Stochastic maximum principle for optimal control of a class of nonlinear spdes with dissipative drift. *SIAM J. Control. Optim.*, 54:341–371, 2016.

[22] Grazieschi, P., Leocata, M., Mascart, C., Chevallier, J., Delarue, F., and Tanré, E. Network of interacting neurons with random synaptic weights. *ESAIM: ProcS*, 65:445–475, 2019.

[23] M. H.A.Davis and G. Burstein. A deterministic approach to stochastic optimal control with application to anticipative control. *Stochastics and Stochastic Reports*, 40(3-4):203–256, 1992.

[24] P. Johnson, J. Pedersen, G. Peskir, and C. Zucca. Detecting the presence of a random drift in brownian motion. *Stochastic Processes and their Applications*, 2021.

[25] R. E. Kalman. The theory of optimal control and the calculus of variations. In R. Bellman, editor, *Mathematical Optimization Techniques*, pages 309–332. University of California Press, 2021.

[26] N. Kazamaki. *Continuous exponential martingales and BMO*. Springer, 2006.

[27] P. L. Lions. Optimal control of diffusion processes and Hamilton–Jacobi–Bellman equations part 2 : viscosity solutions and uniqueness. *Communications in Partial Differential Equations*, 8(11):1229–1276, 1983.

[28] The Mathworks, Inc., Natick, Massachusetts. *MATLAB version 9.10.0.1710957 (R2021a) Update 4*, 2021.

[29] O. Menoukeu-Pamen and L. Tangpi. Maximum principle for stochastic control of SDEs with measurable drifts. *arXiv preprint arXiv:2101.06205*, 2021.

[30] P. A. Meyer. *Probability and potentials*, volume 1318. Blaisdell Publishing Company, 1966.

[31] B. Øksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

[32] B. Øksendal and A. Sulem. A maximum principle for optimal control of stochastic systems with delay, with applications to finance. In *Optimal control and partial differential equations (Paris, 4 December 2000)*, pages 64–79, 2001.

[33] C. Orrieri. A stochastic maximum principle with dissipativity conditions. *Discrete & Continuous Dynamical Systems*, 35(11):5499–5519, 2015.

[34] E. Pardoux and S. Peng. Adapted solution of a backward stochastic differential equation. *Systems & Control Letters*, 14(1):55–61, 1990.

[35] S. Peng. A general stochastic maximum principle for optimal control problems. *Siam Journal on Control and Optimization*, 28:966–979, 1990.

[36] H. Pham. On some recent aspects of stochastic control and their applications. *Probability Surveys*, 2:506 – 549, 2005.

[37] L. Pick, A. Kufner, O. John, and S. Fucík. *Function Spaces*, volume 1. Walter de Gruyter, 2012.

[38] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media, 2013.

[39] W. Rudin. *Real and Complex Analysis, 3rd Ed.* McGraw-Hill, Inc., USA, 1987.

[40] W. Stannat and L. Wessels. Deterministic control of stochastic reaction-diffusion equations. *Evolution Equations & Control Theory*, 2020.

[41] J. Štecha and J. Rathouský. Stochastic maximum principle. *IFAC Proceedings Volumes*, 44(1):4714–4720, 2011.

[42] L. Tonelli. *Fondamenti di Calcolo delle Variazioni I.* Zanichelli, 1921.

[43] W. Whitt. *Stochastic-process limits: an introduction to stochastic-process limits and their application to queues*. Springer Science & Business Media, 2002.

[44] G. M. Wing. *A primer on integral equations of the first kind: the problem of deconvolution and unfolding*. SIAM, 1991.

[45] E. Wong and M. Zakai. On the relation between ordinary and stochastic differential equations. *International Journal of Engineering Science*, 3(2):213–229, 1965.

[46] J. Yong and X. Y. Zhou. *Stochastic controls: Hamiltonian systems and HJB equations*, volume 43. Springer Science & Business Media, 1999.

* Dipartimento di Matematica e Applicazioni "Renato Caccioppoli", Università degli Studi di Napoli Federico II, 80126 Napoli, Italy

† Dipartimento di Matematica "G. Peano", Università degli Studi di Torino, Via Carlo Alberto 10, 10123 Torino, Italy

Email address: `giacomo.ascione@unina.it, giuseppe.donofrio@unito.it`