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

Giacomo Ascione 1 · Giuseppe D’Onofrio 2

Accepted: 6 February 2023 / Published online: 10 April 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

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.

Keywords Stochastic differential equation · Euler–Lagrange equation · Geometric Brownian motion

Mathematics Subject Classification 49J55 · 60H10

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 modify 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 [38]). The first one is called the dynamic programming principle, based on Bellman’s optimality principle [9]: 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 [26, 30]. Let us emphasize that one can also adapt the latter to more complex situation (e.g. [8]). The second approach, instead, is based on a stochastic generalization of Pontryagin’s maximum principle [37]. 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) [36]. 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, 12, 14, 22, 23, 31, 35, 43]. 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 [17] where a stochastic control problem is approximated by a sequence of deterministic control problems, obtaining a Wong–Zakai like [48] 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.

An example in which such approximation is of primary importance is given by the stochastic models of single neurons embedded in a network. Indeed, even if we use a standard Leaky Integrate-and-Fire model (see [41] for a review), embedding a single neuron in a network leads to the presence of a stochastic term (representing the main external input received by the neuron) in the drift of the SDE describing the membrane potential (see, for instance [3]). Such kind of problem, that involves linear SDEs with additive noise, has been addressed in [4]. However, more realistic models make use of multiplicative forms of the noise leading to different SDEs, such as the Cox–Ingersoll–Ross squared model (see, for instance [7, 28]) and the inhomogeneous Geometric Brownian motion model (see, for instance [15]). For this reason, it is useful to extend the approximation results proposed in [4] to the multiplicative noise case. As a first step for such an extension, we still consider the linear case, that is to say the one involving the inhomogeneous Geometric Brownian motion.

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

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

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 [34] to neuronal modeling [20, 24] or quickest detection [25]. Moreover, if \( z(t) \) is itself the solution of an SDE, Eq. (1) plays a role in many systems of equations used in epidemiology, climate models, game theory and others [2, 11, 21].

We are interested in finding the best approximation for a solution of Eq. (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

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

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) 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 \( \tilde{u} \in L^1(0, T) \) that minimizes \( J \). In the literature, to the best of our knowledge, few contributions on purely deterministic controls of stochastic equations are available [4, 42].

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. Let us stress that since we are controlling the drift of an SDE, the natural set of competitors is \( L^1(0, T) \), i.e. the minimal assumption so that the solution of the SDE exists. On the other hand, the main difficulty concerning the existence is the possible lack of coercivity of the cost functional. 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 Sect. 2 we first show some basic properties of the solution of Eq. (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 \( J \). To do this, we make use of needle (or spike) variations, that describe the effect of locally changing the value of the control. This choice is justified by the clear interpretation of the control variation. A different, but direct, approach is also described. In Sect. 4 we prove that, under suitable convexity assumptions, the aforementioned Euler–Lagrange equation is also a sufficient condition. In Sect. 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 Sect. 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 [44].

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

\[(H1) \text{ There exists } p \geq 2 \text{ such that for almost any } t \in [0, T], z(t) \in L^p(\Omega, \mathbb{P}) \text{ and } \int_0^T \mathbb{E}[|z(t)|^p] \frac{d}{dt} dt < + \infty.\]

We denote by \( L^2_p(\Omega, \mathbb{P}; [0, T]) \) the space of \( \mathcal{F}_t \)-adapted processes satisfying (H1). Let us also consider a function \( a \in L^\infty(0, T) \). We focus on the linear SDE

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

(3)

where \( X_0 \in L^2(\Omega, \mathbb{P}) \) is \( \mathcal{F}_0 \)-measurable. In particular the following result holds.

**Proposition 1** For any \( p \geq 2 \), the map \( S_{X_0} : L^2_p(\Omega, \mathbb{P}; [0, T]) \rightarrow L^2_p(\Omega, \mathbb{P}; [0, T]) \), such that the process \( S_{X_0}z \) is solution of (3), is well-defined and it holds

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

(4)
where $A(t) = \int_0^t a(s)ds$ and $G(t)$ is the geometric Brownian motion associated to $W(t)$, i.e.

$$G(t) = e^{W(t) - \frac{1}{2}t}.$$  \hspace{1cm} (5)

**Proof** By a simple adaptation of the proof of [33, Theorem 5.2.1], the SDE (3) admits a unique strong solution in $L^2_2(\Omega, \mathbb{P}; [0, T])$ whenever $z \in L^2_p(\Omega, \mathbb{P}; [0, T])$. One can verify, by applying Itô’s formula, that (4) is solution of (3). \hfill \Box

**Remark 1** The same result holds if we substitute $L^2_p(\Omega, \mathbb{P}; [0, T])$ with $L^1(0, T)$, interpreted as a subspace of degenerate stochastic processes.

As a direct consequence of the previous result, the following corollary can be proved by direct calculations.

**Corollary 1** The solution map $S_{X_0}$ 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 $z_1, \ldots, z_n \in L^2_2(\Omega, \mathbb{P}; [0, T])$ it holds

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

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

$$S_{X_1} z_1 - S_{X_2} z_2 = S_{X_1 - X_2} (z_1 - z_2).$$

In particular, $S_0$ is linear.

Next, we want to underline some properties of the moments of $S_{X_0} z(t)$ when $z$ belongs to a certain Banach space. To do this, we recall the following moment estimate for linear SDEs (see [49, Chapter 3, Lemma 4.2]).

**Lemma 1** 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 $Y_0 \in L^{2\kappa}(\Omega, \mathbb{P})$, $a_1, b_1 \in L^\infty(0, T)$ are functions in $L^\infty(0, T)$ with $M = \max\{|a_1|_{L^\infty}, |b_1|_{L^\infty}\}$ and

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

Then there exists a constant $K(\kappa, M, T) > 0$ such that
sup_{t \in [0, T]} E[|Y(t)|^{2\kappa}] \leq K(\kappa, M, T) \left( E[|Y_0|^{2\kappa}] + \left( \int_0^T E[|a_2(t)|^{2\kappa}] \frac{1}{2} dt \right)^\frac{1}{\kappa} \right) + \left( \int_0^T E[|b_2(t)|^{2\kappa}] \frac{1}{2} dt \right) \frac{1}{\kappa}.

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

Remark 2 Actually, the last statement of the Lemma is a direct consequence of the constructive proof presented in [49, Chapter 3, Lemma 4.2].

Applying the previous Lemma to Eq. (3) we get the following result.

Lemma 2 Let \( z \in L^p_2(\Omega, \mathbb{P}; [0, T]) \) and \( X_0 \in L^p(\Omega, \mathbb{P}) \) for some \( p \geq 2 \). Then it holds \( \sup_{t \in [0, T]} E[|S_{X_0}z(t)|^p] < +\infty \). Moreover, if \( X_0 = 0 \) almost surely and \( u \in L^1(0, T) \), then

\[
\sup_{t \in [0, T]} E[|S_0u(t)|^p] \leq K \left( \frac{p}{2}, M, T \right) \|u\|_{L^1}^p,
\]

where \( M = \|a\|_{L^\infty} \) and \( K \) is defined in Lemma 1.

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 Eq. (5) will play a major role. Let us first recall that, it being a Doleans–Dade exponential (see [27, 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 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 \( t > 0 \), \( \log(G(t)) \sim \mathcal{N}(\frac{1}{2}t, t) \). By using the formula of the moment generating function of a Gaussian random variable, it is easy to show that, for any \( q \geq 0 \),

\[
E[G(t)^q] = e^{q(q-1)t}, \quad t \geq 0,
\]

and the same relation holds for \( G'(t) \). Combining Eq. (6) and Doob’s maximal inequality (see [40, Theorem II.1.7]) we get the following bound on the supremum of \( G \) and \( G' \).

\[\square\]
Lemma 3  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

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

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

where we also used Eq. (6). Arguing in the same way on $G'$ we conclude the proof. □

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 6, 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 Eq. (3), obtained by substituting the possibly stochastic drift with a deterministic function. From now on, we will assume that (H1) is satisfied. Let us first introduce a cost functional

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

where $F$ is a suitable function and $\xi_u = S_0(z - u)$. Let us consider the following assumptions on $F$:
It holds $F(t, \xi) \geq 0$ for any $t \in [0, T]$ and $\xi \in \mathbb{R}$;

$F(t, \xi)$ is twice continuously differentiable in the $\xi$ 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;

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 $u \in L^1(0, T)$ such that

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

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 $X(t) = S_{X_0}z(t)$ 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 $X_u(t) = S_{X_0}u(t)$. By affinity of the solution map, we have that $\xi_u(t) := X(t) - X_u(t) = S_0(z - u)(t)$. 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 $\xi_u$. The growth assumption on $F$ can be justified by means of the following non-triviality result.

**Lemma 4** For any $u \in L^1(0, T)$ it holds $J[u] < +\infty$.

**Proof** We have

$$J[u] = \mathbb{E} \left[ \int_0^T F(t, \xi_u(t))dt \right] \leq \int_0^T L(t) \left( 1 + \mathbb{E}[|\xi_u(t)|^\alpha] \right)dt. \quad (7)$$

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

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

where we used Lemma 2. Plugging the previous inequality in (7) we conclude the proof. \qed

The previous result and hypothesis (H2) guarantee that it makes sense to search for a minimizer (if it exists) of $J[u]$.  

Springer
3 Necessary Optimality Conditions

Now let us focus on necessary optimality conditions. Before going into details, let us introduce some notation. For any \( f \in L^1(0, T) \) we denote the set of its Lebesgue points as \( E_f \) (see [19, Sect. 1.7]). To obtain necessary optimality conditions we need the following property, which is a simple consequence of Hölder’s inequality:

**Proposition 2** 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 \( t \in [0, T] \). Then \( h \in L^1(0, T) \) and \( E_f \subseteq E_h \).

Now we are ready to prove the main result of this section.

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

\[
\int_{t_0}^{T} \mathbb{E} \left[ \frac{\partial F}{\partial \xi} (t, \xi(t)) \frac{G(t)}{G(t_0)} e^{(A(t) - A(t_0))} \right] dt = 0, \quad t_0 \in [0, T].
\] (8)

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

\[
u_\varepsilon(t) = \begin{cases} u & t \in I_\varepsilon \\ \bar{u}(t) & \text{otherwise} \end{cases},
\]

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

\[
g(\varepsilon) = J[u_\varepsilon] = \mathbb{E} \left[ \int_{0}^{T} F(t, \xi_\varepsilon(t)) dt \right].
\]

We want to show that \( g \) is right differentiable in 0.

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

\[
F(t, \xi_\varepsilon(t)) - F(t, \xi_\bar{u}(t)) = -\frac{\partial F}{\partial \xi}(t, \xi(t)) \eta_\varepsilon(t)
\]

\[
+ \int_{0}^{1} \theta \frac{\partial^2 F}{\partial \xi^2}(t, \theta \xi_\varepsilon(t) + (1 - \theta)\xi_\bar{u}(t)) \eta_\varepsilon^2(t) d\theta,
\]
where $\eta_\varepsilon(t) := \xi_\varepsilon(t) - \xi_\varepsilon(t)$, and then

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

By the properties of the solution map $S_0$ given in Corollary 1, we have that

$$
\eta_\varepsilon(t) = S_0(u_\varepsilon - \bar{u})(t) = \begin{cases} 
0 & t \leq t_0 - \frac{\varepsilon}{2} \\
G(t)e^{A(t)} \int_{t_0 - \frac{\varepsilon}{2}}^t e^{-A(s)} [u - \bar{u}(s)] ds & t \in I_\varepsilon \\
G(t)e^{A(t)} \int_{I_\varepsilon} e^{-A(s)} [u - \bar{u}(s)] ds & t \geq t_0 + \frac{\varepsilon}{2}
\end{cases},
$$

and then we can split the integral in (9) as

$$
\frac{g(\varepsilon) - g(0)}{\varepsilon} = -\frac{1}{\varepsilon} \mathbb{E} \left[ \int_{I_\varepsilon} \frac{\partial F}{\partial \xi}(t, \xi_\varepsilon(t)) \eta_\varepsilon(t) dt \right] - \frac{1}{\varepsilon} \mathbb{E} \left[ \int_{t_0 + \frac{\varepsilon}{2}}^T \frac{\partial F}{\partial \xi}(t, \xi_\varepsilon(t)) \eta_\varepsilon(t) dt \right]
\begin{array}{c}
+ \frac{1}{\varepsilon} \mathbb{E} \left[ \int_0^T \int_0^1 \theta \frac{\partial^2 F}{\partial \xi^2}(t, \theta \xi_\varepsilon(t) + (1 - \theta) \xi_\varepsilon(t)) \eta_\varepsilon^2(t) d\theta dt \right]
\end{array}
$$

$$
=: I_1(\varepsilon) + I_2(\varepsilon) + I_3(\varepsilon).
$$

First of all, let us show that $\lim_{\varepsilon \to 0} I_1(\varepsilon) = 0$. To do this, we observe that

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

Bringing back the exponent $\alpha \in (0, p)$ in hypothesis (H4), let us consider any $\tilde{p} \in \left(1, \min \{2, \frac{p}{\alpha}\}\right)$ and let $\tilde{q} > 2$ be its conjugate exponent.

By Hölder’s inequality it holds

$$
\begin{align}
|I_1(\varepsilon)| & \leq \frac{1}{\varepsilon} \int_{I_\varepsilon} \mathbb{E} \left[ \left( \left| \frac{\partial F}{\partial \xi}(t, \xi_\varepsilon(t)) \right|^{\tilde{p}} \right)^{\frac{1}{\tilde{p}}} \left( \mathbb{E} \left[ |\eta_\varepsilon(t)|^{\tilde{q}} \right] \right)^{\frac{1}{\tilde{q}}} dt.
\end{align}
$$

\begin{flushright}
\copyright Springer
\end{flushright}
Using the growth hypothesis (H4) we have

$$
\mathbb{E} \left[ \left| \frac{\partial F}{\partial \xi} (t, \xi_\pi(t)) \right|^\tilde{p} \right] \leq L\tilde{p}(t) \mathbb{E} \left[ (1 + |\xi_\pi(t)|^\alpha)^{\tilde{p}} \right] \\
\leq L\tilde{p}(t)^{\tilde{p}-1} (1 + \mathbb{E}[|\xi_\pi(t)|^\alpha])^{\frac{\tilde{p}}{p}} \\
\leq L\tilde{p}(t)^{2\tilde{p}-1} \left( 1 + (\mathbb{E}[|\xi_\pi(t)|^p])^{\frac{\tilde{p}}{p}} \right) \leq CL\tilde{p}(t),
$$

where we used the convexity of the function $x \mapsto x^{\tilde{p}}$, applied Hölder’s inequality a second time with the exponent $\frac{p}{\alpha} > 1$ and finally used Lemma 2. Plugging inequality (12) into (11) and using Lemma 2 on $\eta_\varepsilon$ we get

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

It is not difficult to see that if $t_0 \in E_{\bar{u}}$, then it is also a Lebesgue point for $|u - \bar{u}(t)|$. Thus, being $t_0 \in E$, we conclude that $\lim_{\varepsilon \to 0} I_1(\varepsilon) = 0$. Now let us show that $\lim_{\varepsilon \to 0} I_3(\varepsilon) = 0$. First recall that by Corollary 1 the process $\theta \xi_\pi + (1 - \theta)\xi_\varepsilon$ is solution of an equation of the form (3), hence we can still use Lemma 2 on it. Thus, arguing exactly as for $I_1$ we have

$$
|I_3(\varepsilon)| \leq \frac{1}{\varepsilon} \int_0^T \int_0^1 \theta \mathbb{E} \left[ \left| \frac{\partial^2 F}{\partial \xi^2} (t, \theta \xi_\pi(t) + (1 - \theta)\xi_\varepsilon(t)) \eta_\varepsilon(t) \right| \right] d\theta dt \\
\leq \frac{1}{\varepsilon} \int_0^T \int_0^1 \theta \left( \mathbb{E} \left[ \left| \frac{\partial^2 F}{\partial \xi^2} (t, \theta \xi_\pi(t) + (1 - \theta)\xi_\varepsilon(t)) \right|^{\tilde{p}} \right] \right)^{\frac{1}{\tilde{p}}} d\theta dt \\
\leq \frac{C}{\varepsilon} \left( \int_{I_\varepsilon} |u - \bar{u}(t)|dt \right)^2 \int_0^T L(t)dt \leq C\varepsilon \left( \frac{1}{\varepsilon} \int_{I_\varepsilon} |u - \bar{u}(t)|dt \right).
$$

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, arguing as before by using Hölder’s inequality, we have

$$
\int_{t_0}^T \mathbb{E} \left[ \frac{\partial F}{\partial \xi} (t, \xi_\pi(t)) \eta_\varepsilon(t) \right] 1_{[t_0, +\frac{T}{2}]}(t) dt \leq C \|u - \bar{u}\|_{L^1} \|L\|_{L^1}.
$$
where, for any Borel set $B \subseteq [0, T]$, $1_B$ is its indicator function. Hence, by Fubini’s theorem, we have

$$I_2(\varepsilon) = -\frac{1}{\varepsilon} \int_0^T \mathbb{E} \left[ \frac{\partial F}{\partial \xi}(t, \xi(t)) \eta_\varepsilon(t) \right] 1_{[t_0^+, T]}(t) dt$$

$$= -\int_0^T \mathbb{E} \left[ 1_{[t_0^+, T]}(t) \frac{\partial F}{\partial \xi}(t, \xi(t)) G(t) e^{A(t)} \frac{1}{\varepsilon} \int_{I_\varepsilon} e^{-A(s)} G(s) [u - \bar{u}(s)] ds \right] dt,$$

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. First, we can suppose $\varepsilon_0$ is small enough to have

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

Thus, we only need to show that the stochastic process on the right-hand side is integrable, which is true once we observe that

$$\int_0^T \mathbb{E} \left[ \left( \sup_{t \in [0, T]} G(t) \right) \left( \sup_{t \in [0, T]} G'(t) \right) \left| \frac{\partial F}{\partial \xi}(t, \xi(t)) \right| \right] dt \leq C,$$

where we used Eq. (12) and Lemma 3.

We are left to show that the integrand in Eq. (14) converges almost everywhere. Recalling that $G$ is almost surely continuous, $t_0 \in E$ is almost surely a Lebesgue point for $e^{-A(t)} G(t) [u - \bar{u}(t)]$ by Proposition 2. Hence we achieve

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

Thus, by Dominated Convergence Theorem, we get

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

In conclusion, from Eq. (10) we have

\[
0 \leq \lim_{\varepsilon \to 0} \frac{g(\varepsilon) - g(0)}{\varepsilon} = -(u - \bar{u}(t_0)) \int_{t_0}^{T} \mathbb{E} \left[ \frac{\partial F}{\partial \xi} (t, \xi(t)) e^{A(t)} A(t_0) \frac{G(t)}{G(t_0)} \right] dt, \tag{17}
\]

where the inequality follows from the fact that 0 is a minimum point for \( g \). The latter inequality can be rewritten as

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

Thus, defining \( H(u) := u \int_{t_0}^{T} \mathbb{E} \left[ \frac{\partial F}{\partial \xi} (t, \xi(t)) e^{A(t)} A(t_0) \frac{G(t)}{G(t_0)} \right] dt \) and observing that \( \bar{u}(t_0) \) is a maximum point for \( H(u) \), we get \( H'(\bar{u}(t_0)) = 0 \), that is to say, being \( t_0 \in E \) arbitrary,

\[
\int_{t_0}^{T} \mathbb{E} \left[ \frac{\partial F}{\partial \xi} (t, \xi(t)) e^{A(t)} A(t_0) \frac{G(t)}{G(t_0)} \right] dt = 0, \forall t_0 \in E. \tag{18}
\]

To extend \( E \) to the whole interval \([0, T]\), we show that

\[
t_0 \in [0, T] \to \int_{t_0}^{T} \mathbb{E} \left[ \frac{\partial F}{\partial \xi} (t, \xi(t)) e^{A(t)} A(t_0) \frac{G(t)}{G(t_0)} \right] dt, \tag{19}
\]

is a continuous function. Hence, consider \( t_1 \in [0, T] \) and \( t_2 = t_1 + \delta \) for some \( \delta \) 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

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

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

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

\[
:= I_4(\delta) + I_5(\delta).
\]

Concerning \( I_4(\delta) \), since the exact same argument we adopted for \( I_2(\varepsilon) \) leads to

\[
I_4(\delta) = \int_{t_1}^{t_1+\delta} \mathbb{E} \left[ \left| \frac{\partial F}{\partial \xi}(t, \xi_\pi(t))e^{A(t) - A(t_1)}G(t) \right| \right] dt \leq C \int_{t_1}^{t_1+\delta} L(t),
\]

we have \( \lim_{\delta \to 0} I_4(\delta) = 0 \). Concerning \( I_5(\delta) \), we can use dominated convergence theorem after observing that, for \( i = 1, 2 \), as in Eq. (15),

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

where the right-hand side is independent of \( t_i \) and integrable. Thus, by almost surely continuity of \( G \), we get that \( \lim_{\delta \to 0} I_5(\delta) = 0 \). This is enough to conclude the proof.

\[\square\]

**Remark 3** Needle variations describe the effect of altering the value of the control in a small interval, giving a way to look at the effects on the trajectory. For further details, see, for instance, [29, Remark 4.14]. Furthermore, we could use any variation of the form \( u + \varepsilon v \) for any \( v \in L^1(0, T) \) and show that \( J \) is Gateaux differentiable (see [10]) in any \( u \in L^1(0, T) \) with Gateaux derivative given by

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

and then Equation (8) can be restated as \( \partial_u J[v] = 0, \ \forall v \in L^1(0, T) \). This would require similar steps and lead to the same necessary condition. From this point of view, such equation is a consequence of Fermat’s theorem applied directly on \( J \) and then we can refer to it 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 \( J \).
However, a clearer description of the action of the control variation on the trajectory is indeed obtained by adopting needle variations. Finally, it is clear that the same result holds for local minimizers.

4 Sufficient Optimality Conditions

In this section we want to investigate whether Eq. (8) is also a sufficient condition. 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 2** Suppose Hypotheses (H1) to (H5) are satisfied. Let \( \bar{u} \in L^1(0, T) \) be a solution of Eq. (8). Then \( \bar{u} \) is a global minimizer of \( J \).

**Proof** Let us consider \( \bar{u} \in L^1(0, T) \) solution of Eq. (8) 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

\[
F(t, \xi_u(t)) - F(t, \xi_{\bar{u}}(t)) \geq \frac{\partial F}{\partial \xi}(t, \xi_{\bar{u}}(t))(\xi_u(t) - \xi_{\bar{u}}(t))
\]

\[
= \int_0^t \frac{\partial F}{\partial \xi}(t, \xi_{\bar{u}}(t)) e^{\lambda(t-A(s))} \frac{G(t)}{G(s)} \bar{u}(s) - u(s) ds,
\]

and then

\[
J[u] - J[\bar{u}] \geq \int_0^T \mathbb{E} \left[ \int_0^t \frac{\partial F}{\partial \xi}(t, \xi_{\bar{u}}(t)) e^{\lambda(t-A(s))} \frac{G(t)}{G(s)} \bar{u}(s) - u(s) ds \right] dt. \tag{20}
\]

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

\[
\mathbb{E} \left[ \int_0^T \frac{\partial F}{\partial \xi}(t, \xi_{\bar{u}}(t)) e^{\lambda(t-A(s))} \frac{G(t)}{G(s)} \bar{u}(s) - u(s) ds \right] \leq C \| \bar{u} - u \|_1 \mathbb{E} \left[ \sup_{t \in (0, T)} \left| \frac{\partial F}{\partial \xi}(t, \xi_{\bar{u}}(t)) \right| \left( \sup_{\tau \in (0, T)} G(\tau) \right) \right]. \tag{21}
\]
One can show that the right hand side of Inequality (21) is integrable by arguing in the same way as for Inequality (16). Hence, we can use Fubini’s theorem in Eq. (20) to achieve

\[ J[u] - J[\bar{u}] \geq \int_0^T (\bar{u}(s) - u(s)) \int_0^T \mathbb{E} \left[ \frac{\partial F}{\partial \xi}(t, \xi_\bar{u}(t)) e^{A(t) - A(s)} \frac{G(t)}{G(s)} \right] dt ds = 0, \quad (22) \]

\( \bar{u} \) being a solution of (8). The fact that \( u \in L^1 \) is arbitrary concludes the proof. \( \square \)

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

**Proposition 3** 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

\[
J[\theta u_1 + (1 - \theta) u_2] = \mathbb{E} \left[ \int_0^T F(t, S_0(z - (\theta u_1 + (1 - \theta) u_2))(t)) dt \right] \\
= \mathbb{E} \left[ \int_0^T F(t, \theta S_0(z - u_1)(t) + (1 - \theta) S_0(z - u_2)(t)) dt \right] \\
\leq \theta \mathbb{E} \left[ \int_0^T F(t, S_0(z - u_1)(t)) dt \right] + (1 - \theta) \mathbb{E} \left[ \int_0^T F(t, S_0(z - u_2)(t)) dt \right] \\
= \theta J[u_1] + (1 - \theta) J[u_2],
\]

where we used hypothesis (H5) and the third statement of Corollary 1. 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 \( \Omega \) such that

I. \( \mathbb{P}(\Omega) > 0 \);  
II. for any \( \omega \in \Omega \), \( S_0(z - u_i)(\cdot, \omega) \) is continuous in \([0, T]\) for \( i = 1, 2 \);  
III. for any \( \omega \in \Omega \), there exists \( t(\omega) \) such that \( S_0(z - u_1)(t(\omega), \omega) \neq S_0(z - u_2)(t(\omega), \omega) \).

In particular, combining II and III we have that for any \( \omega \in \Omega \) there exists an interval \( I(\omega) \) such that \( S_0(z - u_1)(t, \omega) \neq S_0(z - u_2)(t, \omega) \) for any \( t \in I(\omega) \). Strict convexity
then follows from the fact that the inequality (23) becomes strict due to hypothesis (H5+) as one considers the contribution of the expectation on $\Omega_1$ of the integral over the random interval $I(\cdot)$. \hfill $\square$

**Remark 4** Let us observe that Theorem 2 can be seen as a direct consequence of Proposition 3 and Remark 3, by using the inequality

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

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

Combining Theorem 2 with Proposition 3 we obtain the following Corollary.

**Corollary 2** Suppose Hypotheses (H1) to (H5+) are satisfied. Then Eq. (8) admits at most one solution.

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 4** Let hypotheses (H1) to (H4) hold. Then $J : L^1(0, T) \to \mathbb{R}$ is continuous.

**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

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

(24)
where we used hypothesis (H3). Let us consider $p$ as in (H1) and $\alpha$ as in (H4) and, let $\tilde{p} \in (1, \min \{2, \frac{p}{\alpha}\})$ and $\tilde{q}$ such that $\frac{1}{\tilde{p}} + \frac{1}{\tilde{q}} = 1$. By Hölder’s inequality we have

$$\mathbb{E} \left[ \left| \frac{\partial F}{\partial \xi}(t, \theta \xi u_2(t) + (1 - \theta)\xi u_1(t)) \right| \right] \leq \mathbb{E} \left[ \left| \frac{\partial F}{\partial \xi}(t, \theta \xi u_2(t) + (1 - \theta)\xi u_1(t)) \right|^{\tilde{p}} \right]^{\frac{1}{\tilde{p}}} \mathbb{E} \left[ (|\xi u_2(t) - \xi u_1(t)|^{\tilde{q}}) \right]^{\frac{1}{\tilde{q}}}.$$  \hspace{1cm} (25)

Let us consider the first term. Let us suppose that $r < 1$. Observe, by Corollary 1, that $\theta \xi u_2(t) + (1 - \theta)\xi u_1(t) = S_0(z - \theta u_2 - (1 - \theta)u_1)$, hence we can use Lemma 2 on it. Moreover, from Lemma 1 and the fact that $\|u_2\|_{L^1} \leq 1 + \|u_1\|_{L^1}$, it is clear that we can consider an upper bound that only depends on $\|u_1\|_{L^1}$. Thus we have, by hypothesis (H4) and arguing as for (12):

$$\mathbb{E} \left[ \left| \frac{\partial F}{\partial \xi}(t, \theta \xi u_2(t) + (1 - \theta)\xi u_1(t)) \right|^{\tilde{p}} \right]^{\frac{1}{\tilde{p}}} \leq C (\|u_1\|_{L^1}) L(t).$$  \hspace{1cm} (26)

Concerning the second term, again Corollary 1 and Lemma 2 give us

$$\mathbb{E} \left[ (|\xi u_2(t) - \xi u_1(t)|^{\tilde{q}}) \right]^{\frac{1}{\tilde{q}}} = \mathbb{E} \left[ (|S_0(u_2 - u_1)(t)|) \right]^{\frac{1}{\tilde{q}}} \leq C \|u_2 - u_1\|_{L^1(0, T)} < Cr.$$  \hspace{1cm} (27)

Combining Eqs. (26) and (27) in Equation (25) and integrating, we get

$$\left| J[u_2] - J[u_1] \right| < C \left( \|u_1\|_{L^1} \right) r.$$

A suitable choice of $r < 1$ concludes the proof.

\[\square\]

**Remark 5** If, moreover, Hypothesis (H5) holds, then $J$ is also weakly lower semi-continuous in $L^1$. This is a direct consequence of Mazur’s theorem [16, Theorem 3.9].

Let us introduce the set of simple functions:

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

where $B([0, T])$ is the Borel $\sigma$-algebra on the interval $[0, T]$. As a direct consequence of the density of $\mathcal{S}$ in $L^1$ and Proposition 4, we have the following Corollary.

**Corollary 3** Let hypotheses (H1) to (H4) hold. Then

$$\inf_{u \in L^1(0, T)} J[u] = \inf_{u \in \mathcal{S}} J[u].$$
Now we want to penalize our functional $J$ to obtain a coercive functional. To do this, let us first give the following definition.

Definition 1 A function $\Psi : [0, +\infty) \to [0, +\infty)$ is called a Young function (see [39]) if there exists a right-continuous non-decreasing function $\psi : [0, +\infty) \to [0, +\infty)$ such that $\psi(s) = 0$ if and only if $s = 0$, $\lim_{s \to +\infty} \psi(s) = +\infty$ and

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

Young functions satisfy different important properties. Here we recall some of them (see [39, Lemma 4.2.2]).

Lemma 5 Any Young function $\Psi$ 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 6 The previous lemma also tells us that for any Young function $\Psi$, $u \in \mathbb{R} \mapsto \Psi(|u|)$ is differentiable with derivative 0 for $u = 0$.

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

$$J_{\delta, \Psi} : u \in L^1(0, T) \mapsto J[u] + \delta F_{\Psi} [u] \in \mathbb{R} \cup \{+\infty\}, \text{ where } F_{\Psi} [u] = \int_0^T \Psi(|u(t)|)dt.$$  

Remark 7 There exist some $u \in L^1(0, T)$ such that $J_{\delta, \Psi} [u] = +\infty$ (see, for instance, [39, Remark 4.2.4]). In particular we can define the Orlicz class

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

and observe that $J_{\delta, \Psi} [u] < +\infty$ if and only if $u \in L^\Psi(0, T)$. Let us stress that $L^\Psi(0, T)$ is in general not a vector space. See [39, Chapter 4] for further details.

First of all, we want to show that for any $\delta > 0$ and any Young function $\Psi$, the functional $J_{\delta, \Psi}$ admits a minimum. To do this, we recall the following result, which is classical in Calculus of Variations (see [18, 45]).

Proposition 5 For any Young function $\Psi$, the functional $F_{\Psi} : L^1(0, T) \mapsto \mathbb{R} \cup \{+\infty\}$ is weakly lower semicontinuous.

Now we are ready to show that $J_{\delta, \Psi}$ admits a minimum.
Theorem 3 Let $\delta > 0$ and $\Psi$ be a Young function. Suppose hypotheses (H1) to (H5) hold. Then there exists a function $\bar{u}_{\delta, \Psi} \in L^\Psi(0, T)$ 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 $\Psi$ 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 $\delta F_{\Psi}$, thus it is convex. Moreover, $J$ is continuous by Proposition 4 and then, in particular, lower semicontinuous, while $\delta F_{\Psi}$ is lower semicontinuous by Proposition 5. Thus $J_{\delta, \Psi}$ is convex and lower semicontinuous and then weakly lower semicontinuous by a direct application of Mazur’s theorem [16, Theorem 3.9]. Moreover, by de la Vallée-Poussin theorem (see [32, Theorem T22]) and Dunford–Pettis theorem (see [13, Theorem 4.30]) we know that the function $J_{\delta, \Psi}$ is coercive in the weak topology of $L^1(0, T)$. Existence of the minimum follow from the direct methods of Calculus of Variations.

\[\square\]

Remark 8 Let us stress that $L^1(0, T)$ is a natural choice of admissible controls. It might seem that the main difficulty in the solution of the minimization problem is due to the fact that $L^1(0, T)$ is not reflexive. However, if we consider the same problem in $L^q(0, T)$ for $q > 1$, then we still do not have coercivity of the cost functional with respect to the weak topology, as, in general, the functional cannot control the $L^q$-norm. Indeed, as a consequence of hypotheses (H1) to (H4) and Lemma 1, we know that there exists a constant $C > 0$ such that

$$J[u] \leq C(1 + \|u\|^q_{L^1(0, T)}), \quad \forall u \in L^1(0, T),$$

hence there exists a sequence $\{u_n\}_{n \in \mathbb{N}} \subset L^q(0, T)$ (one can choose, for instance, $u_n(t) = \min\{t^{-\frac{1}{q}} n\}$) such that $\|u_n\|_{L^q(0, T)} \to +\infty$ but $J[u_n] \leq M$ for some constant $M > 0$ independent of $n \in \mathbb{N}$. Furthermore, in this case, we cannot choose a completely arbitrary penalization: for instance if we consider $\Psi(t) = \frac{2t^{q+1}}{q+1}$, then, despite $\Psi$ is a Young function, the penalized functional $J_{\delta, \Psi}$ is still not coercive in the weak topology of $L^q(0, T)$, while it is so in the weak topology of $L^1(0, T)$.

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

Theorem 4 Let hypotheses (H1) to (H5) hold and consider a Young function $\Psi$. The family of functions $\{\bar{u}_{\delta, \Psi}\}_{\delta > 0}$ defined in Theorem 3 constitute a minimizing family for $J$.

Proof Set $m = \inf_{u \in L^1(0, T)} J[u]$ and let us recall that $\mathfrak{S} \subset L^\Psi(0, T)$. Consider any $u \in \mathfrak{S}$ and observe that

$$m \leq J[\bar{u}_{\delta, \Psi}] \leq J_{\delta, \Psi}[\bar{u}_{\delta, \Psi}] \leq J_{\delta, \Psi}[u] = J[u] + \delta F_{\Psi}[u].$$
Taking the limit superior and inferior as $\delta \to 0$ and then the infimum on $S$, we get, by Corollary 3
\[ m \leq \liminf_{\delta \to 0} J[\bar{u}_\delta, \psi] \leq \limsup_{\delta \to 0} J[\bar{u}_\delta, \psi] \leq m. \]
\[ \square \]

The 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** Let hypotheses (H1) to (H5) hold and suppose there exist two Young functions $\Psi_i$, $i = 1, 2$, and two constants $C, \delta_0 > 0$ such that $\mathcal{F}_{\Psi_i}[\bar{u}_\delta, \psi_1] \leq C$ for any $\delta \in (0, \delta_0)$, where the functions $\bar{u}_\delta, \psi_1$ are defined in Theorem 3. Then there exists $\bar{u} \in L^1(0, T)$ such that
\[ \inf_{u \in L^1(0, T)} J[u] = J[\bar{u}], \]
and $\bar{u}_{\delta_n}, \psi_1 \to \bar{u}$ for a sequence $\delta_n \downarrow 0$. Moreover, if hypothesis (H5+) holds, then $\bar{u}$ is unique.

**Proof** Let us consider any sequence $\delta_n \downarrow 0$ with $\delta_1 < \delta_0$. The hypothesis on $\mathcal{F}_{\Psi_2}$ implies, via the de la Vallée-Poussin theorem, that the sequence $\{\bar{u}_{\delta_n}, \psi_1\}_{n \in \mathbb{N}}$ is uniformly integrable and then the Dunford–Pettis theorem ensures that it is weakly relatively compact. Hence, there exists $\bar{u}$ such that $\bar{u}_{\delta_n}, \psi_1 \to \bar{u}$ for a non relabelled subsequence $\delta_n$. By Theorem 4 we know that $\lim_{n} J[\bar{u}_{\delta_n}, \psi_1] = \inf_{u \in L^1(0, T)} J[u] =: m$.

On the other hand, being $J$ convex and continuous in $L^1(0, T)$, we know that it is weakly lower semicontinuous and then
\[ m = \lim_{n} J[\bar{u}_{\delta_n}, \psi_1] \geq J[\bar{u}] \geq m, \]
thus $J[\bar{u}] = m$, concluding the first part of the proof. The second statement follows from the strict convexity of $J$. \[ \square \]

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

**Corollary 4** Let hypotheses (H1) to (H5) hold and consider a Young function $\Psi$. Suppose there exists $\bar{u} \in L^{\Psi}(0, T)$ such that $J[\bar{u}] = \min_{u \in L^1(0, T)} J[u]$. Then, the family $\{\bar{u}_\delta, \psi\}_{\delta > 0}$ defined in Theorem 3 satisfies the hypotheses of Theorem 5 with $\psi_1 = \psi_2 = \Psi$. Moreover, if hypothesis (H5+) holds, then $\bar{u}_{\delta_n}, \psi_1 \to \bar{u}$ for some sequence $\delta_n \to 0$.

**Proof** Recall that $\bar{u}_\delta, \psi$ is a global minimizer of $J_\delta, \psi$ and $\bar{u}$ is a global minimizer of $J$, thus we have
\[ J[\bar{u}_\delta, \psi] + \delta \mathcal{F}_\psi[\bar{u}_\delta, \psi] = J_\delta, \psi[\bar{u}_\delta, \psi] \leq J_\delta, \psi[\bar{u}] = J[\bar{u}] + \delta \mathcal{F}_\psi[\bar{u}] \leq J[\bar{u}_\delta, \psi] + \delta \mathcal{F}_\psi[\bar{u}], \]
that implies
\[ F_{\Psi}[\bar{u}_{\ delta}, \Psi] \leq F_{\Psi}[\bar{u}] =: C, \]
concluding the proof. \( \square \)

**Remark 9** By de la Vallée-Poussin theorem we know that, if there exists a minimizer \( \bar{u} \in L^1(0, T) \) for \( J \), there always exists a Young function \( \Psi \) such that \( \bar{u} \in L^\Psi(0, T) \). Such Young function can be used to apply Corollary 4.

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

**Theorem 6** Let hypotheses (H1) to (H5+) hold and suppose \( J \) admits a global minimizer \( \bar{u} \in L^{\tilde{p}}(0, T) \) for some \( \tilde{p} > 4 \) and consider \( \Psi(t) = C_0 t^\tilde{p} \). Let \( \{\bar{u}_{\delta}, \Psi\}_{\delta > 0} \) be defined in Theorem 3. Let \( X_0 \in L^{\tilde{p}}(\Omega, \mathbb{P}) \). Then there exists a sequence \( \delta_n \to 0 \) such that \( S_{X_0} \bar{u}_{\delta_n}, \Psi \to S_{X_0} \bar{u} \) in \( C([0, T]) \) in distribution.

**Proof** Without loss of generality we suppose \( C_0 = 1 \). Observe that, by Corollary 1 and Eq. (4),
\[
S_{X_0} \bar{u}_{\delta}, \Psi(t) - S_{X_0} \bar{u}(t) = \left| S_0(\bar{u}_{\delta}, \Psi - \bar{u})(t) \right| = \left| e^{-A(t)} \int_0^t \frac{e^{-A(s)}}{G(s)} \bar{u}_{\delta}, \Psi(s) ds - \int_0^t \frac{e^{-A(s)}}{G(s)} \bar{u}(s) ds \right|.
\]

Consider \( \delta_n \to 0 \) as in Corollary 4, so that \( \bar{u}_{\delta_n}, \Psi \to \bar{u} \) in \( L^1(0, T) \). Since \( \frac{e^{-A(\cdot)}}{G(\cdot)} \) is almost surely a continuous function, we have that
\[
\lim_{n \to +\infty} \left| S_{X_0} \bar{u}_{\delta_n}, \Psi(t) - S_{X_0} \bar{u}(t) \right| = 0, \tag{28}
\]
almost surely and for any fixed \( t \in [0, T] \). This is enough to guarantee the convergence in any finite-dimensional distribution. To extend the convergence in distribution to the whole paths, we need to show that the sequence \( S_{X_0} \bar{u}_{\delta_n}, \Psi \) is tight. Let us denote, for simplicity, \( X_\delta = S_{X_0} \bar{u}_{\delta_n}, \Psi \). Consider \( 0 \leq t_1 < t_2 \leq T \) and observe that, by Eq. (4), it holds
\[
X_\delta(t_2) - X_\delta(t_1) = \left( G(t_2)e^{A(t_2)} - G(t_1)e^{A(t_1)} \right) \left( X_0 + \int_{t_1}^{t_2} \frac{e^{-A(s)}}{G(s)} \bar{u}_{\delta_n}, \Psi(s) ds \right)
+ G(t_2)e^{A(t_2)} \int_{t_1}^{t_2} \frac{e^{-A(s)}}{G(s)} \bar{u}_{\delta_n}, \Psi(s) ds.
\]
Observe that \( \hat{p} \frac{2}{p} > 2 \) and \( \frac{4\hat{p}}{4 + \hat{p}} > 2 \), then we can choose \( \overline{p} \in \left( 2, \min \left\{ \frac{4\hat{p}}{4 + \hat{p}}, \frac{\hat{p}}{2} \right\} \right) \). This technical choice will be clarified later. By convexity inequality,

\[
\mathbb{E}[|X_\delta(t_2) - X_\delta(t_1)|^{\overline{p}}] 
\leq 2^{\overline{p} - 1} \left( \mathbb{E} \left[ |G(t_2)e^{\overline{A}(t_2)} - G(t_1)e^{\overline{A}(t_1)}|^{\overline{p}} \right] X_0 + \int_0^{t_1} \frac{e^{-\overline{A}(s)}}{G(s)} \overline{\mu}_\delta,\psi(s) ds \right)^{\overline{p}} 
+ \mathbb{E} \left[ G^{\overline{p}}(t_2)e^{\overline{p}A(t_2)} \left| \int_{t_1}^{t_2} \frac{e^{-\overline{A}(s)}}{G(s)} \overline{\mu}_\delta,\psi(s) ds \right|^{\overline{p}} \right) = 2^{\overline{p} - 1}(I_1 + I_2).
\]

(29)

Let us first work with \( I_2 \). Using Hölder’s inequality with exponent 2 and Jensen’s inequality we achieve

\[
I_2 \leq \mathbb{E}[G^{2\overline{p}}(t_2)e^{2\overline{p}A(t_2)}]^{\frac{1}{2}}(t_2 - t_1)\overline{p} \mathbb{E} \left[ \int_{t_1}^{t_2} \frac{e^{-\overline{A}(s)}}{G^{2\overline{p}}(s)} |\overline{\mu}_\delta,\psi(s)|^{2\overline{p}} ds \right]^{\frac{1}{2}} \leq C_1(T, \overline{p})(t_2 - t_1)^{\gamma_1} \mathbb{E} \left[ \int_{t_1}^{t_2} \frac{e^{-2\overline{p}A(s)}}{G^{2\overline{p}}(s)} |\overline{\mu}_\delta,\psi(s)|^{2\overline{p}} ds \right]^{\frac{1}{2}},
\]

where we set \( \gamma_1 = \overline{p} - \frac{1}{2} \), with \( \gamma_1 > 1 \) by the choice of \( \overline{p} \), and we used Lemma 3 to control \( \mathbb{E}[G^{2\overline{p}}(t_1)e^{2\overline{p}A(t_1)}]^{\frac{1}{2}} \). Again, by Lemma 3, we have

\[
\mathbb{E} \left[ \int_{t_1}^{t_2} \frac{e^{-2\overline{p}A(s)}}{G^{2\overline{p}}(s)} |\overline{\mu}_\delta,\psi(s)|^{2\overline{p}} ds \right]^{\frac{1}{2}} \leq C_2(T, \overline{p}) \left( \int_0^T |\overline{\mu}_\delta,\psi(s)|^{2\overline{p}}(s) ds \right)^{\frac{1}{2}}. 
\]

(31)

Observing that, by definition of \( \overline{p} \), it holds \( \frac{\overline{p} - 2\overline{p}}{2\overline{p}} > 1 \), we can use it as exponent in Hölder’s inequality, obtaining

\[
\left( \int_0^T |\overline{\mu}_\delta,\psi|^{2\overline{p}}(s) ds \right)^{\frac{\overline{p}}{2}} \leq \left( \int_0^T |\overline{\mu}_\delta,\psi|^{\overline{p}}(s) ds \right)^{\frac{\overline{p}}{2}} T^{\frac{\overline{p} - 2\overline{p}}{2\overline{p}}} \leq \left( \int_0^T |\overline{\mu}_\delta|^{\overline{p}}(s) ds \right)^{\frac{\overline{p}}{2}} T^{\frac{\overline{p} - 2\overline{p}}{2\overline{p}}},
\]

 Springer
where we also used the fact that, as in Corollary 4, \( \mathcal{F}_\Psi[\overline{\mu}_\delta, \Psi] \leq \mathcal{F}_\Psi[\overline{\mu}] \). Plugging last inequality in (31) we achieve

\[
\mathbb{E} \left[ \int_{t_1}^{t_2} \frac{e^{-2\overline{\mu}A(s)}}{G^{2\overline{\mu}}(s)} |\overline{\mu}_\delta, \Psi(s)|^{2\overline{\mu}} ds \right]^{\frac{1}{2}} \leq C_3(T, \overline{\rho}, \overline{p}, \overline{\mu})
\]

and then we obtain, from Eq. (30),

\[
I_2 \leq C_4(T, \overline{\rho}, \overline{p}, \overline{\mu})(t_2 - t_1)^{\gamma_1}. \quad (32)
\]

Now let us consider \( I_1 \). Observe that \( \frac{4\overline{\rho}}{\overline{p} + \overline{\rho}} < 4 \) and then also \( \overline{p} < 4 \). By Hölder’s inequality with exponent \( \frac{4}{\overline{p}} > 1 \) we get

\[
I_1 \leq \mathbb{E} \left[ |G(t_2)e^{A(t_2)} - G(t_1)e^{A(t_1)}|^{4\overline{p}} \right]^{\frac{4}{\overline{p}}} \mathbb{E} \left[ X_0 + \int_0^{t_1} \frac{e^{-A(s)}}{G(s)} \overline{\mu}_\delta, \Psi(s) ds \right]^{\frac{4}{\overline{p}}}.
\]

Recalling that \( \frac{4\overline{p}}{\overline{p} + \overline{\rho}} > 1 \), since \( \overline{p} > 2 > \frac{4}{\overline{p}} \), we can use the convexity inequality, Jensen’s inequality and Lemma 3 to get

\[
\mathbb{E} \left[ X_0 + \int_0^{t_1} \frac{e^{-A(s)}}{G(s)} \overline{\mu}_\delta, \Psi(s) ds \right]^{\frac{4}{\overline{p}}} \leq 2^{\frac{4}{\overline{p}} - 1} \left( \mathbb{E} \left[ |X_0|^{4\overline{p}} \right] + \mathbb{E} \left[ \int_0^{t_1} \frac{e^{-A(s)}}{G(s)} |\overline{\mu}_\delta, \Psi(s)|^{4\overline{p}} ds \right] \right)^{\frac{4}{\overline{p}}}
\]

\[
\leq 2^{\frac{4}{\overline{p}} - 1} \left( \mathbb{E} \left[ |X_0|^{4\overline{p}} \right] + t_1^{\frac{4}{\overline{p}} - 1} \mathbb{E} \left[ \int_0^{t_1} \frac{e^{-A(s)}}{G^{4\overline{p}}(s)} |\overline{\mu}_\delta, \Psi(s)|^{4\overline{p}} ds \right] \right)
\]

\[
\leq C_5(T, \overline{p}) \left( \mathbb{E} \left[ |X_0|^{4\overline{p}} \right] + \int_0^{T} |\overline{\mu}_\delta, \Psi(s)|^{4\overline{p}} ds \right).
\]

Moreover, being \( \overline{p} < \frac{4\overline{p}}{\overline{p} + \overline{\rho}} \), it holds \( \frac{\overline{p}(4\overline{p})}{4\overline{p}} > 1 \) and we can use it as exponent for Hölder’s inequality, obtaining,

\[
\mathbb{E} \left[ |X_0|^{\frac{4\overline{p}}{\overline{p} + \overline{\rho}}} \right] \leq \left( \mathbb{E}[|X_0|^4] \right)^{\frac{\overline{p}(4\overline{p})}{4\overline{p}}} < +\infty
\]
and, as before,
\[
\int_0^T |\tilde{u}_{\delta, \psi}(s)|^{\frac{4p}{4-mp}} ds \leq \left( \int_0^T |\tilde{u}(s)|^{\frac{4p}{mp}} ds \right)^{\frac{4p}{mp}} T^{1-\frac{4p}{mp}},
\]
so that Inequality (33) becomes
\[
I_1 \leq C_6(T, \tilde{p}, \tilde{p}, \tilde{u}, X_0) \mathbb{E}\left[ |G(t_2)e^{A(t_2)} - G(t_1)e^{A(t_1)}|^4 \right]^{\frac{2}{p}}. \tag{34}
\]

To estimate the remaining part of \(I_1\), define \(Y(t) = G(t)e^{A(t)}\) and observe, by Itô’s formula, that \(Y(t)\) solves \(dY(t) = a(t) Y(t) dt + Y(t) dW(t)\) in \([t_1, t_2]\). Setting then \(\tilde{Y}(t) = Y(t + t_1) - \gamma Y(t_1)\) for \(t \in [0, t_1 - t_2]\) and \(\tilde{Y}(t) = Y(t_2) - \gamma Y(t_1)\) for \(t \in [t_2 - t_1, T]\), we know that the following stochastic differential relation holds:
\[
d\tilde{Y}(t) = a(t + t_1) \mathbf{1}_{[0, t_1 - t_2]}(t) (\tilde{Y}(t) + \gamma Y(t_1)) dt + (\tilde{Y}(t) + \gamma Y(t_1)) \mathbf{1}_{[0, t_2 - t_1]}d\tilde{W}(t),
\]
where \(\tilde{W}(t) = W(t + t_1) - W(t_1)\) is still a Brownian motion. Setting \(\tilde{M} = \|a\|_{L^\infty(0, T)}\), we have, by Lemma 1 with \(K = K(2, M, T) \geq K(2, M, t_2 - t_1)\),
\[
\mathbb{E}[|\tilde{Y}(t_2 - t_1)|^4] \leq \sup_{t \in [0, t_2 - t_1]} \mathbb{E}[|\tilde{Y}(t_1)|^4] \leq K \left( \int_0^{t_2 - t_1} \mathbb{E}[|a(t + t_1) Y(t_1)|^4]^{\frac{1}{4}} dt \right)^4 + \left( \int_0^{t_2 - t_1} \mathbb{E}[|Y(t_1)|^4]^{\frac{1}{2}} dt \right)^2 \leq C_7(T) (t_2 - t_1)^2.
\]

Recalling the definition of \(\tilde{Y}\) and plugging last inequality in (34) we get
\[
I_1 \leq C_8(T, \tilde{p}, \tilde{p}, \tilde{u}, X_0) \mathbb{E}[|X(t_2) - X(t_1)|^{2\gamma}], \tag{35}
\]
where \(\gamma = \frac{8}{9} > 1\). Now set \(\gamma = \min\{\gamma_1, \gamma_2\} > 1\) and combine inequalities (32) and (35) with (29) to conclude that
\[
\mathbb{E}[|X_\delta(t_2) - X_\delta(t_1)|^{\frac{4p}{mp}}] \leq C_9(T, \tilde{p}, \tilde{p}, \tilde{u}, X_0) (t_2 - t_1)^{\gamma p}.
\]

This, together with the fact that \(X_\delta(0) = X_0\) for any \(\delta > 0\), guarantees that \(\{X_\delta\}_{\delta > 0}\) is tight (see, for instance, [46, Theorem 11.6.5]). Thus, by a corollary of Prohorov’s theorem (see, for instance, [46, Corollary 11.6.2]), we know that \(X_{\delta_n} \rightarrow S_{X_0} \tilde{u}\) in \(C([0, T])\), concluding the proof. \(\square\)

Theorem 6 guarantees that even if \(\tilde{u}_{\delta, \psi}\) does not converge strongly to \(\tilde{u}\) (due, for instance, to a highly oscillatory behaviour), it can be still used to approximate the
process $S_{X_0 \overline{u}}$. This comes in handy in the application context, whenever one has to numerically determine some functional properties of $S_{X_0 \overline{u}}$.

Let us finally 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 7** Let hypotheses (H1) to (H5) hold and $\Psi$ be a Young function with strictly increasing continuous derivative $\psi$. Then $\overline{u}_{\delta, \psi}$ is the unique solution of

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

(36)

**Proof** First observe that being $\overline{u}_{\delta, \psi} \in L_{\psi}(0, T)$, then $\Psi(|u_{\delta, \psi}|) \in L^1(0, T)$. 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 $(0, T)$, $E_{\psi}$ the set of Lebesgue points of $\Psi(|\overline{u}_{\delta, \psi}|)$ in $(0, T)$ and $E = E_{\overline{u}} \cap E_L \cap E_{\psi}$. Let $t_0 \in E$, fix a real number $u \in \mathbb{R}$ and $\varepsilon_0 > 0$ small enough to have $(t_0 - \frac{\varepsilon_0}{2}, t_0 + \frac{\varepsilon_0}{2}) \subset (0, T)$. Define, for any $\varepsilon \in (0, \varepsilon_0)$, $I_{\varepsilon} := \left( t_0 - \frac{\varepsilon}{2}, t_0 + \frac{\varepsilon}{2} \right)$, $u_{\varepsilon}$ accordingly as in Theorem 1, $g_1(\varepsilon) := J_{\delta, \psi}[u_{\varepsilon}]$, $g_2(\varepsilon) := J[u_{\varepsilon}]$ and $g_3(\varepsilon) := \mathcal{F}_\psi[u_{\varepsilon}]$, so that $g_1 = g_2 + \delta g_3$ and

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

Let us only study the second incremental ratio. It holds, since $t_0 \in E \subset E_{\psi}$,

$$
\lim_{\varepsilon \to 0} \frac{g_3(\varepsilon) - g_3(0)}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{I_{\varepsilon}} (\Psi(|u|) - \Psi(|\overline{u}_{\delta, \psi}(t)|)) dt = \Psi(|u|) - \Psi(|\overline{u}_{\delta, \psi}(t_0)|).
$$

Combining the last equality with Eq. (17) we obtain

$$
0 \leq \lim_{\varepsilon \to 0} \frac{g_1(\varepsilon) - g_1(0)}{\varepsilon} = \delta \Psi(|u|) - \delta \Psi(|\overline{u}_{\delta, \psi}(t_0)|)
$$

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

(37)

where we used the notation introduced in Theorem 1 and the fact that 0 is a minimum point of $g_1$. Setting

$$
H(u) := -\delta \Psi(|u|) + u \int_{t_0}^{T} \mathbb{E} \left[ \frac{\partial F}{\partial \xi} (t, \xi_{\overline{u}_{\delta, \psi}}(t)) e^{A(t) - A(t_0)} \frac{G(t)}{G(t_0)} \right] dt,
$$
inequality (37) implies that \( \max_{u \in \mathbb{R}} H(u) = H(\bar{u}_\delta, \psi(t_0)) \) and then, differentiating \( H \) in \( \bar{u}_\delta, \psi(t_0) \),

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

Finally, the right-hand side being continuous (as we have shown in the proof of Theorem 1), we achieve Eq. (36).

To show that the latter equation is also a sufficient condition, let us just observe that, for any \( u_{\delta, \psi} \in \mathcal{L}_\psi(0, T) \),

\[
\mathcal{F}_\psi[u] - \mathcal{F}_\psi[\bar{u}_\delta, \psi] \geq \int_0^T \frac{\bar{u}_\delta, \psi(s)}{|\bar{u}_\delta, \psi(s)|} \psi(\bar{u}_\delta, \psi(s))(u(s) - \bar{u}_\delta, \psi(s)) ds,
\]

where we used the fact that \( \Psi(|x|) \) is convex with derivative \( \frac{x}{|x|} \psi(|x|) \). Combining (38) and (22) and arguing as in Theorem 2 we conclude the proof. \( \square \)

**Remark 10** Let us stress, as we did for \( J \), that \( \mathcal{F}_\psi \) is Gateaux-differentiable in \( \mathcal{L}_\psi(0, T) \) with Gateaux derivative

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

One can show that the previous quantity is finite for any \( v \in \mathcal{L}_\psi(0, T) \) by means of Hölder’s inequality for Orlicz spaces \([39, \text{Theorem 4.7.5}]\). As a consequence, we obtain that also \( J_{\delta, \psi}[u] \) is Gateaux-differentiable in \( \mathcal{L}_\psi(0, T) \) with Gateaux derivative

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

and Eq. (36) follows from Fermat’s theorem. For this reason, we can refer to (36) as the Euler–Lagrange equation for \( J_{\delta, \psi} \). Let us emphasize that the function \( H(u) \) defined in the proof of Theorem 7 is, in some sense, the Hamiltonian function associated to \( J_{\delta, \psi} \).

With a suitable choice of the Young function \( \Psi \), we can guarantee continuity for \( \bar{u}_\delta, \psi \). Indeed the next Corollary directly follows from Eq. (36).

**Corollary 5** Let hypotheses (H1) to (H5) hold and set \( \psi(t) = \frac{t^{2n}}{2n} \) for any \( n = 1, 2, 3, \ldots \). Then \( \bar{u}_\delta, \psi \) is continuous.

\( \square \) Springer
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, the following Proposition can be easily proved by direct calculations.

**Proposition 6** Fix $p \geq 2$ and let $z \in \mathcal{L}_p^2(\Omega, \mathbb{P}; [0, T])$ for some $\tilde{p} > p$. Then $J^{(p)}$ satisfies hypotheses (H1) to (H5+).

The Euler–Lagrange equation (8) for the functional $J^{(p)}$ can be stated as

$$\int_0^T \mathbb{E} \left[ |\xi_{\tau}(\tau)|^{p-2} \xi_{\tau}(\tau) e^{A(\tau)-A(t)} \frac{G(\tau)}{G(t)} \right] d\tau = 0, \ \forall t \in [0, T].$$

(39)

To introduce a penalization on such functional $J^{(p)}$, let us consider $\Psi^{(2)}(x) = \frac{x^2}{2}$, $\mathcal{F}^{(2)} := \mathcal{F}_{\Psi^{(2)}}$ and $J^{(p,2)} := J^{(p)} + \delta \mathcal{F}^{(2)}$. Its Euler–Lagrange equation (36) can be recast as

$$\bar{u}_\delta(t) = \frac{1}{\delta} \int_0^T \mathbb{E} \left[ |\xi_{\tau}(\tau)|^{p-2} \xi_{\tau}(\tau) e^{A(\tau)-A(t)} \frac{G(\tau)}{G(t)} \right] d\tau, \ \forall t \in [0, T].$$

(40)

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. On the other hand, when $p = 2$, we are able to restate both Eqs. (39) and (40) 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, and let us consider $J^{(2)}$ as before. By Proposition 6 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 (39). Actually, in this case, we can restate the equation as a Fredholm integral equation of the first kind.

**Proposition 7** Let \( z \in L^2_p(\Omega, \mathbb{P}; [0, T]) \) for some \( p > 2 \). Then \( \bar{u} \in L^1(0, T) \) is the unique minimizer of \( J^{(2)} \) if and only if

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

where

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

and

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

**Proof** We already know that \( \bar{u} \) is the unique minimizer of \( J^{(2)} \) if and only if it solves

\[
\int_0^T \mathbb{E} \left[ \int_0^t \frac{G^2(t)}{G(t_0)G(s)} e^{2A(t) - A(t_0) - A(s)} (z(s) - \bar{u}(s)) ds \right] dt = 0, \quad \forall t_0 \in [0, T], \quad (44)
\]

that is Eq. (39) for \( p = 2 \) after we substitute \( \xi_{\bar{u}} \) with the explicit formula given in Eq. (4). 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 observe that, by using Cauchy–Schwartz inequality, Lemma 3 and the fact that \( z \in L^2_p(\Omega, \mathbb{P}; [0, T]) \),

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

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

\[
\leq C \mathbb{E} \left[ \left( \sup_{s \in [0, T]} G'(s) \right)^4 \left( \sup_{s \in [0, T]} G(s) \right)^4 \right] \mathbb{E} \left[ \left( \int_0^T |z(s)|^2 ds + \| \bar{u} \|_{L^1}^2 \right) \right] \leq C.
\]
Hence, we can use Fubini’s theorem to rewrite Eq. (44) as

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

By using the definition of $G(t)$ as in (5) it is clear that $\frac{G^2(t)}{G(t_0)G(s)}$ is a lognormal random variable with

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

Substituting last equality in Eq. (45) and setting $k(t, s; a)$ and $\mathcal{Z}(t)$ as in Eqs. (42) and (43) we conclude the proof.

An analogous result can be shown for Eq. (40).

**Proposition 8** Let $z \in L^2_p(\Omega, \mathbb{P}; [0, T])$ for some $p > 2$. Then $\bar{u}_\delta \in L^1(0, T)$ is the unique minimizer of $J_\delta^{(2,2)}$ if and only if

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

where $k$ and $\mathcal{Z}$ are defined in Eqs. (42) and (43).

We omit the proof since it is identical to the previous one.

Both Propositions 7 and 8 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 9** Let $z \in L^2_p(\Omega, \mathbb{P}; [0, T])$ for some $p > 2$ be independent of the Brownian motion $W(t)$. Then $J^{(2)}$ admits as unique minimizer the function $\bar{u}(t) = \mathbb{E}[z(t)].$
Moreover, it holds

\[ J^{(2)}[\bar{u}] = \frac{1}{2} \int_0^T \int_0^t e^{2A(t) - A(s) - A(\tau) + t - \max\{s, \tau\}} \text{Cov}(z(s), z(\tau)) ds d\tau dt, \quad (47) \]

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

**Proof** By Proposition 7 we know that \( \bar{u} \) is the unique minimizer of \( J^{(2)} \) if and only if it solves Eq. (41). Thus, we only have to show that \( \bar{u}(t) = \mathbb{E}[z(t)] \) solves that equation. This is clear since, being \( z \) independent of \( W \), we get by Eq. (43)

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

Now let us evaluate \( J^{(2)}[\bar{u}] \). We have, by Eq. (4),

\[ J^{(2)}[\bar{u}] = \frac{1}{2} \int_0^T \mathbb{E} \left[ G^2(t) e^{2A(t)} \left( \int_0^t e^{-A(s)} \frac{G(s)}{G(\tau)} (z(s) - \bar{u}(s)) d\tau \right)^2 \right] dt \]

\[ = \frac{1}{2} \int_0^T \mathbb{E} \left[ \int_0^t \int_0^t G^2(t) e^{2A(t) - A(s) - A(\tau)} \frac{(z(s) - \bar{u}(s))(z(\tau) - \bar{u}(\tau))}{G(s)G(\tau)} ds d\tau \right] dt. \]

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, let us observe that, by the independence of \( z \) and \( W \) (and thus of \( z \) and \( G \)) and Lemma 3 we have

\[ \int_0^T \mathbb{E} \left[ \int_0^T \int_0^t G^2(t) e^{2A(t) - A(s) - A(\tau)} \frac{|z(s) - \bar{u}(s)||z(\tau) - \bar{u}(\tau)|}{G(s)G(\tau)} ds d\tau \right] dt \]

\[ \leq C \mathbb{E} \left[ \left( \sup_{s \in [0, T]} G(s) \right)^2 \left( \sup_{s \in [0, T]} G'(s) \right)^2 \right] \int_0^T \int_0^T \mathbb{E}[|z(s) - \bar{u}(s)||z(\tau) - \bar{u}(\tau)|] ds d\tau \]

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

where in the second last inequality we used the Cauchy–Schwartz inequality in the expectation and then Jensen’s inequality while in the last one we just observed that \( \mathbb{E}[|z(s) - \bar{u}(s)|^2] \leq \mathbb{E}[|z(s)|^2], \) being \( \bar{u}(s) = \mathbb{E}[z(s)]. \) Hence we can use Fubini’s theorem and the fact that \( z \) is independent of \( W \) in Eq. (48) to conclude the proof. \( \square \)
Fig. 1 Numerical solutions of Eq. (46) with $T = 1$, $Z(t)$ given in Eq. (50) and $k(t, s; a)$ given in Eq. (49), for different values of $\delta$. Precisely, reading left-to-right top-to-bottom we have $\overline{u}_\delta$ for $\delta = 10^{-n}$ with $n = 1, \ldots, 6$. $N$ is fixed to 100, so that we have 601 nodes for each $\overline{u}_\delta$.

Remark 11 The 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 $u(t) = E[z(t)] \equiv 1$. Let $a(t) \equiv -1$, and then $A(t) = -t$, and $T = 1$. In this case we have

$$k(t, s; -1) = e^{-|t-s|} - e^{-1+\min[t,s]}, \quad (49)$$

and

$$Z(t) = e^{-1-t}(-e + e^{t} - 3e^{2t} + 2e^{1+t} + e^{2t}). \quad (50)$$

Let us also stress that, since $\text{Cov}(z(s), z(\tau)) = e^{\min[\tau,s]} - 1$, we have, by Eq. (47),

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

For any $\delta = 10^{-n}$, let $\overline{u}_n$ be the solution of Eq. (46). To obtain a numerical evaluation of $\overline{u}_n$ we used Nyström method (see [6]), where the quadrature formula is a composite 7-th order closed Newton–Cotes formula on a subdivision of $[0, 1]$ by $N$ subintervals, since we expect a highly oscillatory behaviour for small values of $\delta$. Such solutions are visualized in Fig. 1.

Evidently, $\overline{u}_n$ does not strongly converge to 1 as $\delta \to 0$. To show a numerical evidence that $\overline{u}_n \to \overline{u}$, we 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 1.

We can also numerically show that $J^{(2)}[\overline{u}_n] \to J^{(2)}[\overline{u}]$. In particular, we simulate $\xi_{\overline{u}_n}$ by means of Euler’s scheme (see [5]), we integrate $\xi_{\overline{u}_n}^2$ by using a quadrature formula and then we evaluate $J^{(2)}[\overline{u}_n]$ by a Monte–Carlo approach. Let us underline that
Table 1 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 \( \int_{0}^{1} t^{j} \, dt = \frac{1}{j} \)

| \( n \) | \( j = 0 \) | \( j = 1 \) | \( j = 2 \) | \( j = 3 \) |
|-------|-------|-------|-------|-------|
| 1     | 0.6141 | 0.2535 | 0.1431 | 0.0930 |
| 2     | 0.8878 | 0.4067 | 0.2495 | 0.1734 |
| 3     | 0.9670 | 0.4690 | 0.3034 | 0.2211 |
| 4     | 0.9899 | 0.4901 | 0.3235 | 0.2403 |
| 5     | 0.9968 | 0.4968 | 0.3301 | 0.2483 |
| 6     | 0.9988 | 0.4988 | 0.3321 | 0.2488 |
| 7     | 0.9991 | 0.4991 | 0.3325 | 0.2491 |
| 8     | 0.9992 | 0.4992 | 0.3325 | 0.2492 |
| 9     | 0.9992 | 0.4992 | 0.3325 | 0.2492 |

\( \overline{u}_{n} \) is fixed to 100 and the values of the integrals are obtained by using the same quadrature formula as applied before to determine \( \overline{u}_{n} \).
Table 2  Numerically estimated values of $J^{(2)}[\overline{u}_n]$ for $n = 1, \ldots, 5$, in comparison with $J^{(2)}[\overline{u}]$

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

$N$ is fixed to 100, and 100,000 trajectories are simulated for each numerical estimate.

by Jensen’s inequality, Lemma 2 and the fact that $\overline{u}_n \rightarrow \overline{u}$, we know that $\int_0^1 \overline{\xi}_n(t)dt$ admits uniformly bounded second moments and then the Monte–Carlo approach converges due to the Strong Law of Large Numbers. While, on one hand, the convergence $J^{(2)}[\overline{u}_n] \rightarrow J^{(2)}[\overline{u}]$ is justified by Theorem 4, 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 $\xi$. The estimated values of $J^{(2)}[\overline{u}_n]$ for $n = 1, \ldots, 5$ are given in Table 2.

With this example, we want to highlight the fact that even if the minimizer of $J^{(2)}$ is known and quite regular, the solutions of the penalized problem converge towards it 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 $S_{X_0}\overline{u}$, that, despite the weak convergence of $\overline{u}_\delta$ towards $\overline{u}$, is in the overall approximated well enough by $S_{X_0}\overline{u}_\delta$, as shown in Theorem 6.

As already stated in Remark 11, Proposition 9 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 9. Thus, let us first determine (at least numerically) the solutions $\overline{u}_n$, with $\delta = 10^{-n}$ of the penalized problem. According to Eqs. (42) and (43) we have

\[
k(t, s; 1) = \frac{1}{3} \left( e^{3-\min\{t,s\}}-2 \max\{t,s\} - e^{t-s} \right), \quad (51)
\]

\[
Z(t) = \frac{1}{6} \left( 3 - 2e^t + e^{2-2t}(2e - 3) \right). \quad (52)
\]

As before, we can exploit some numerical solutions $\overline{u}_n$ of Eq. (46) by using Nyström’s method, as shown in Fig. 2.

To have a qualitative idea on whether a solution of Eq. (41) 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 $p > 1$. This numerical evidence lets us conjecture that $\overline{u}_n$ is uniformly bounded in $L^p(0, 1)$.
Table 3  Numerically estimated values of $\|u_n(t)\|_{L^p(0,1)}^p$ 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 $\pi_n$.

|     | $n = 1$ | $n = 2$ | $n = 3$ | $n = 4$ | $n = 5$ | $n = 6$ | $n = 7$ | $n = 8$ | $n = 9$ |
|-----|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| $p = 2$ | 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   |
for some $p > 1$ and then, by Theorem 5, that a minimizer $\bar{u}$ of $J^{(2)}$ exists. Let us also conjecture, with the help of Table 3, that $\bar{u} \in L^2(0, 1)$.

With this idea in mind, let us evaluate numerically the solution of Eq. (41). 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 use a Galërkin-type method (see [47, Sect. 6.3]) based on the orthogonal system of Legendre polynomials in $L^2(0, 1)$. The approximation of $\bar{u}$ up to the 5-th degree polynomial is shown in Fig. 3. To have a numerical evidence of the fact that $\bar{u}_n \to \bar{u}$, we compare $\int_0^1 t^j \bar{u}_n(t) \, dt$ with $\int_0^1 t^j \bar{u}(t) \, dt$ for $j = 0, 1, 2, 3$ in Table 4.

**Remark 12** Let us emphasize that one must pay attention to the choice of the numerical method to solve Eq. (46). Indeed, one cannot exclude a priori an highly oscillatory behaviour of the solution of (46), as shown 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.
Table 4 Numerically estimated values of $\int_0^1 t^j \pi_n(t) dt$ for $n = 1, \ldots, 9$ and $j = 0, 1, 2, 3$, in comparison with $\int_0^1 t^j \pi(t) dt$

|       | $n = 1$ | $n = 2$ | $n = 3$ | $n = 4$ | $n = 5$ | $n = 6$ | $n = 7$ | $n = 8$ | $n = 9$ | $\int_0^1 t^j \pi(t) dt$ |
|-------|---------|---------|---------|---------|---------|---------|---------|---------|---------|-----------------|
| $j = 0$ | 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          |

$N$ is fixed to 100 and the values of the integrals are obtained by using the same quadrature formula as applied before to determine $\pi_n$.
Table 5 Numerically estimated values of $J^{(2)}[\bar{\pi}_n]$ for $n = 1, \ldots, 3$, in comparison with $J^{(2)}[\bar{\pi}]$ and $J^{(2)}[\mathbb{E}[z(\cdot)]]$

|                | $J^{(2)}[\bar{\pi}_1]$ | $J^{(2)}[\bar{\pi}_2]$ | $J^{(2)}[\bar{\pi}_3]$ | $J^{(2)}[\bar{\pi}]$ | $J^{(2)}[\mathbb{E}[z(\cdot)]]$ |
|----------------|-------------------------|-------------------------|-------------------------|-----------------------|----------------------------------|
|                | 0.0836                  | 0.0572                  | 0.0510                  | 0.0577                | 0.1505                           |

$N$ is fixed to 100, and 100,000 trajectories are simulated for each numerical estimate. 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 $J^{(2)}[\bar{\pi}]$ 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 $J^{(2)}[\bar{\pi}_n]$ to the best error.

Just taking a quick look at Fig. 3, it seems clear that $\bar{u} \neq 1 \equiv \mathbb{E}[z(\cdot)]$. Indeed, in this case, $\mathbb{E}[z(\cdot)]$ is not a minimizer of $J^{(2)}$, since

$$\int_0^1 k(t, s; 1) ds = -\frac{1}{6} e^{-3t} (e^3 - 4e^{3t} + 2e^{4t} - 2e^{3+t} + 3e^{1+2t}) \neq Z(t),$$

so that $\mathbb{E}[z(\cdot)]$ does not solve Eq. (41). Actually, the expected value seems to be quite far from the optimal approximation. This can be observed by evaluating $J^{(2)}[\bar{\pi}_n]$, $J^{(2)}[\bar{\pi}]$ and $J^{(2)}[\mathbb{E}[z(\cdot)]]$ as before. Again, the evaluations of $J^{(2)}[\bar{\pi}_n]$ for big values of $n$ are not reliable due to the stiffness of the underlying problem. The results are exposed in Table 5: here it is evident that the expected value is not the optimal approximation.

**Remark 13** 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.

**Acknowledgements** The authors would like to thank the anonymous referee for their precious comments.

**Funding** This research is partially supported by MIUR-PRIN 2017, project Stochastic Models for Complex Systems, No. 2017JFFHSH and by Gruppo Nazionale per l’Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA-INdAM).

**Declarations**

**Competing interests** The authors have not disclosed any competing interests.

**References**

1. Andersson, D., Djehiche, B.: A maximum principle for SDEs of mean-field type. Appl. Math. Optim. 63, 341–356 (2011)
2. Arnold, L.: Random Dynamical Systems. Monographs in Mathematics. Springer, New York (1998)
3. Ascione, G., Carfora, M.F., Pirozzi, E.: A stochastic model for interacting neurons in the olfactory bulb. Biosystems 185, 104030 (2019)
4. Ascione, G., D’Onofrio, G., Kostal, L., Pirozzi, E.: An optimal Gauss–Markov approximation for a process with stochastic drift and applications. Stochastic Process. Appl. 130(11), 6481–6514 (2020)
5. Asmussen, S., Glynn, P.W.: Stochastic Simulation: Algorithms and Analysis, vol. 57. Springer, New York (2007)
6. Atkinson, K.E.: The Numerical Solution of Integral Equations of the Second Kind, vol. 4. Cambridge University Press, Cambridge (1997)
7. Bauermann, J., Lindner, B.: Multiplicative noise is beneficial for the transmission of sensory signals in simple neuron models. Biosystems 178, 25–31 (2019)
8. Bayraktar, E., Cosso, A., Pham, H.: Randomized dynamic programming principle and Feynman–Kac representation for optimal control of McKean–Vlasov dynamics. Trans. Am. Math. Soc. 370(3), 2115–2160 (2018)
9. Bellman, R.: Dynamic programming and stochastic control processes. Inf. Control 1(3), 228–239 (1958). https://doi.org/10.1016/S0019-9958(58)80003-0
10. Berger, M.S.: Nonlinearity and Functional Analysis: Lectures on Nonlinear Problems in Mathematical Analysis, vol. 74. Academic Press, Cambridge (1977)
11. Bhattacharya, R., Majumdar, M.: Random Dynamical Systems: Theory and Applications. Cambridge University Press, Cambridge (2007)
12. Bonnans, J.F., Silva, F.J.: First and second order necessary conditions for stochastic optimal control problems. Appl. Math. Optim. 65, 403–439 (2012)
13. Brezis, H.: Functional Analysis, Sobolev Spaces and Partial Differential Equations. Springer, Berlin (2010)
14. Cordoni, F., Di Persio, L.: A maximum principle for a stochastic control problem with multiple random terminal times. Math. Eng. 2, 557 (2020). https://doi.org/10.3934/mine.2020025
15. D’Onofrio, G., Lansky, P., Pirozzi, E.: On two diffusion neuronal models with multiplicative noise: The mean first-passage time properties. Chaos Interdiscipl. J. Nonlinear Sci. 28(4), 043103 (2018)
16. Dacorogna, B.: Direct Methods in the Calculus of Variations, vol. 78. Springer, New York (2007)
17. Davis, M.H.A., Burstein, G.: A deterministic approach to stochastic optimal control with application to anticipative control. Stochast. Stochast. Rep. 40(3–4), 203–256 (1992). https://doi.org/10.1080/17442509208833790
18. De Giorgi, E.: Semicontinuity Theorems in the Calculus of Variations. Quaderni dell’Accademia Pontaniana, Accademia Pontaniana, Berlin (2008)
19. Evans, L.C., Gariepy, R.F.: Measure Theory and Fine Properties of Functions. CRC Press, Cambridge (2015)
20. Faugeras, O., Soret, E., Tanré, E.: Asymptotic behaviour of a network of neurons with random linear interactions. Preprint at http://arxiv.org/abs/Probability (2019)
21. Flandoli, F., Tonello, E.: An Introduction to Random Dynamical Systems for Climate (2019). https://courseclimath19.sciencesconf.org/data/pages/RDS_4_Climate_Flandoli_v2.pdf. Accessed 07 Oct 2021
22. Fuhrman, M., Orrieri, C.: Stochastic maximum principle for optimal control of a class of nonlinear spdes with dissipative drift. SIAM J. Control. Optim. 54, 341–371 (2016)
23. Fuhrman, M., Hu, Y., Tessitore, G.: Stochastic maximum principle for optimal control of SPDEs. Preprint at http://arXiv.org/abs/1302.0286 (2013)
24. Grazieschi, P., Leocata, M., Mascart, C., Chevallier, J., Delarue, F., Tanré, E.: Network of interacting neurons with random synaptic weights. ESAIM ProcS 65, 445–475 (2019). https://doi.org/10.1051/proc/201965445
25. Johnson, P., Pedersen, J., Peskir, G., Zucca, C.: Detecting the presence of a random drift in Brownian motion. Stochast. Process. Appl. (2021). https://doi.org/10.1016/j.spa.2021.05.006
26. Kalman, R.E.: The theory of optimal control and the calculus of variations. In: Bellman, R. (ed.) Mathematical Optimization Techniques, pp. 309–332. University of California Press, Berkeley (2021)
27. Kazamaki, N.: Continuous Exponential Martingales and BMO. Springer, Berlin (2006)
28. Lánský, P., Lanska, V.: Diffusion approximation of the neuronal model with synaptic reversal potentials. Biol. Cybern. 56(1), 19–26 (1987)
29. Lewis, A.D.: The Maximum Principle of Pontryagin in Control and in Optimal Control. Handouts for the Course Taught at the Universitat Politecnica de Catalunya (2006)
30. Lions, P.L.: Optimal control of diffusion processes and Hamilton–Jacobi–Bellman equations part 2: viscosity solutions and uniqueness. Commun. Part. Differ. Equ. 8(11), 1229–1276 (1983). https://doi.org/10.1080/03605308308820301
31. Menoukeu-Pamen, O., Tangpi, L.: Maximum principle for stochastic control of SDEs with measurable drifts. Preprint at http://arxiv.org/abs/2101.06205 (2021)
32. Meyer, P.A.: Probability and Potentials, vol. 1318. Blaisdell Publishing Company, Waltham (1966)
33. Øksendal, B.: Stochastic Differential Equations: An Introduction with Applications. Springer, New York (2013)
34. Øksendal, B., Sulem, A.: 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), pp. 64–79 (2001)
35. Orrieri, C.: A stochastic maximum principle with dissipativity conditions. Discret. Contin. Dyn. Syst. 35(11), 5499–5519 (2015)
36. Pardoux, E., Peng, S.: Adapted solution of a backward stochastic differential equation. Syst. Control Lett. 14(1), 55–61 (1990). https://doi.org/10.1016/0167-6911(90)90082-6
37. Peng, S.: A general stochastic maximum principle for optimal control problems. SIAM J. Control. Optim. 28, 966–979 (1990)
38. Pham, H.: On some recent aspects of stochastic control and their applications. Probab. Surv. 2, 506–549 (2005). https://doi.org/10.1214/154957805100000195
39. Pick, L., Kufner, A., John, O., Fučík, S.: Function Spaces, vol. 1. Walter de Gruyter, Berlin (2012)
40. Revuz, D., Yor, M.: Continuous Martingales and Brownian Motion, vol. 293. Springer, New York (2013)
41. Sacerdote, L., Giraudo, M.T.: Stochastic integrate and fire models: a review on mathematical methods and their applications. In: Stochastic Biomathematical Models, pp. 99–148 (2013)
42. Stannat, W., Wessels, L.: Deterministic control of stochastic reaction–diffusion equations. Evol. Equ. Control Theory (2020). https://doi.org/10.3934/eect.2020087
43. Štecha, J., Rathouský, J.: Stochastic maximum principle. IFAC Proc. 44(1), 4714–4720 (2011). https://doi.org/10.3182/20110828-6-IT-1002.01501
44. The Mathworks, Inc., Natick, Massachusetts: MATLAB version 9.10.0.1710957 (R2021a) Update 4 (2021)
45. Tonelli, L.: Fondamenti di Calcolo delle Variazioni I. Zanichelli (1921)
46. Whitt, W.: Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues. Springer, New York (2002)
47. Wing, G.M.: A Primer on Integral Equations of the First Kind: The Problem of Deconvolution and Unfolding. SIAM, Philadelphia (1991)
48. Wong, E., Zakai, M.: On the relation between ordinary and stochastic differential equations. Int. J. Eng. Sci. 3(2), 213–229 (1965). https://doi.org/10.1016/0020-7225(65)90045-5
49. Yong, J., Zhou, X.Y.: Stochastic Controls: Hamiltonian Systems and HJB Equations, vol. 43. Springer, New York (1999)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.