

A Certified BiLSTM System for Reflected BSDEs with Application to Energy Storage Valuation

Hani Abidi

Keywords and phrases: Reflected backward stochastic differential equation; optimal stopping; deep learning; bidirectional LSTM; a posteriori certification; energy storage valuation; penalization method; neural network approximation.

This manuscript generalizes the results presented in the *International Conference on Recent Advances in Mathematics and Informatics (ICRAMI 2025)* (Tunisia, November 2025). The present paper extends the analysis from standard BSDEs to reflected BSDEs, introduces a residual-based certification framework with rigorous error bounds, and develops a unified error decomposition for penalization and discretization errors.

Abstract A certified numerical framework is developed for reflected backward stochastic differential equations driven by Brownian motion. The approach combines penalization, an implicit Euler discretization, and a data-driven approximation of the discrete scheme.

The main result is a residual-based *a posteriori* stability estimate: any adapted pair that approximately satisfies the penalized scheme is close to the exact discrete solution in the $\mathbb{S}^2(\pi)$ and $\mathbb{H}^2(\pi)$ norms. This yields an ideal certification functional in terms of conditional expectations and a computable version via regression, up to a conditional-expectation approximation error.

A unified error decomposition is established, separating penalization, discretization, and approximation errors with explicit dependence on the parameters. The framework provides rigorous accuracy guarantees for learned approximations in high dimension.

1 Introduction

Backward stochastic differential equations (BSDEs), introduced in [1], play a fundamental role in stochastic control, mathematical finance, and the probabilistic representation of nonlinear partial differential equations. Reflected BSDEs (RBSDEs), developed in [2], extend this framework by incorporating a constraint enforced through an increasing process. They provide a probabilistic formulation of optimal stopping problems and obstacle-type partial differential equations, with well-established results on existence, uniqueness, and connections to Snell envelopes and variational inequalities; see also [3, 4].

Classical references on stochastic control, BSDE methods, and martingale techniques underlying these formulations include [5, 6, 7, 8].

Related applications involving switching and impulse control have also received significant attention; see [9, 10, 11]. These settings further highlight the need for reliable numerical methods in complex stochastic environments.

The numerical approximation of BSDEs and their reflected counterparts has been extensively studied. Time discretization schemes were analyzed in [12], and discrete-time Monte Carlo methods were developed in [13, 14, 15]. These approaches extend naturally to penalized formulations of reflected problems. However, several difficulties persist: the stiffness induced by the penalization parameter, reduced accuracy near the free boundary, and the absence of computable error bounds for individual numerical realizations.

Recent advances in machine learning have led to efficient numerical methods for high dimensional BSDEs and parabolic PDEs; see [16, 17, 18, 19]. These approaches demonstrate strong empirical performance in high dimension, where traditional grid-based methods become infeasible. However, they typically rely on training losses as indirect error indicators and do not provide rigorous guarantees for individual approximations.

For sequence models, recurrent neural networks such as Long Short-Term Memory (LSTM) architectures [20] and their bidirectional extensions [21] offer flexible parametrizations for time-dependent problems and have recently attracted attention for high-dimensional stochastic approximation problems. Nevertheless, ensuring adaptedness and providing quantitative error control remains challenging.

This manuscript generalizes the results presented in [22], which appeared in the *International Conference on Recent Advances in Mathematics and Informatics (ICRAMI 2025)* (Tunisia, November 2025). The present paper extends the analysis from standard BSDEs to reflected BSDEs, introduces a residual-based certification framework with rigorous error bounds, and develops a unified error decomposition for penalization and discretization errors.

To address these challenges, we develop a certified numerical framework for RBSDEs. We consider a penalized formulation of the reflected equation combined with an implicit time discretization scheme.

For the approximation of the resulting discrete scheme, we employ a neural-network parametrization based on a bidirectional Long Short-Term Memory (BiLSTM) architecture. The construction is designed so that all predictions remain adapted, while non-causal information is used only during training. This ensures consistency with the underlying stochastic structure.

At the certification level, we establish a residual-based *a posteriori* stability result showing that any adapted process that approximately satisfies the discrete equations is quantitatively close to the exact discrete penalized solution in the $\mathbb{S}^2(\pi)$ and $\mathbb{H}^2(\pi)$ norms. This yields a computable certification criterion expressed solely in terms of discrete quantities.

Building on this result, we derive a unified error decomposition that separates penalization, time-discretization, and approximation errors. This provides an explicit and interpretable bound on the total numerical error.

The certification results are established at the level of the discrete-time penalized BSDE. Convergence to the continuous-time reflected solution follows from these bounds together with classical results on penalization and time discretization.

For clarity, we restrict attention to a fixed exogenous operational policy and the associated optimal stopping problem, although extensions to joint control–stopping formulations are standard.

The paper is organized as follows. Section 2 introduces the model and its RBSDE formulation. Section 3 presents the penalized problem. Section 4 analyzes the time discretization. Section 5 describes the neural approximation. Section 6 develops the certification framework. Section 7 establishes the main error decomposition result. Numerical experiments are reported in Section 8, and concluding remarks are given in Section 9.

2 Problem Formulation: Energy Storage as a Reflected BSDE

We formulate the energy storage valuation problem under uncertainty and derive its representation as a reflected backward stochastic differential equation (RBSDE). The framework combines stochastic price dynamics, constrained storage evolution, and an optimal stopping feature.

2.1 Mathematical Framework

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$ be a filtered probability space satisfying the usual conditions and supporting a standard Brownian motion $(W_t)_{t \in [0, T]}$.

For an adapted continuous process U , define

$$\|U\|_{\mathbb{S}^2}^2 := \mathbb{E} \left[\sup_{t \in [0, T]} |U_t|^2 \right],$$

and denote by $\mathbb{S}^2$ the set of such processes with finite norm.

For a predictable process V , define

$$\|V\|_{\mathbb{H}^2}^2 := \mathbb{E} \left[\int_0^T |V_t|^2 dt \right],$$

and denote by $\mathbb{H}^2$ the corresponding space. Let $\mathbb{A}^2$ be the set of nondecreasing adapted continuous processes K with $K_0 = 0$ and $K_T \in L^2$.

On a uniform grid $\pi = \{t_k = k\Delta t\}_{k=0}^N$ with $\Delta t := T/N$, we define the discrete norms

$$\|U\|_{\mathbb{S}^2(\pi)}^2 := \mathbb{E} \left[\max_{0 \leq k \leq N} |U_{t_k}|^2 \right], \quad \|V\|_{\mathbb{H}^2(\pi)}^2 := \sum_{k=0}^{N-1} \mathbb{E}[|V_{t_k}|^2] \Delta t. \quad (2.1)$$

2.2 Model and Admissible Policies

The logarithm of the electricity spot price is modeled by the Ornstein–Uhlenbeck process

$$dX_t = \kappa(\theta - X_t) dt + \sigma dW_t, \quad X_0 = x_0, \quad (2.2)$$

with parameters $\kappa > 0$, $\theta \in \mathbb{R}$, and $\sigma > 0$. The spot price is $P_t = e^{X_t}$.

The storage level evolves according to

$$dE_t = u_t dt, \quad E_0 \in [0, E_{\max}], \quad (2.3)$$

where u_t denotes the power exchange rate. The admissible set is constrained by physical bounds $u_{\min} < 0 < u_{\max}$.

On the grid π , a control $(u_{t_k})_{k=0}^{N-1}$ is called admissible if it is $(\mathcal{F}_{t_k})$ -adapted and satisfies

$$u_{t_k} \in [\underline{u}(E_{t_k}), \bar{u}(E_{t_k})], \quad \text{a.s.}, \quad (2.4)$$

where

$$\underline{u}(E) := \max(u_{\min}, -E/\Delta t), \quad \bar{u}(E) := \min(u_{\max}, (E_{\max} - E)/\Delta t). \quad (2.5)$$

These constraints ensure that the storage process remains in $[0, E_{\max}]$ almost surely.

In practice, admissibility is enforced by projection:

$$u_{t_k} = \Pi_{[\underline{u}(E_{t_k}), \bar{u}(E_{t_k})]}(u_{t_k}^{\text{raw}}), \quad \Pi_{[a,b]}(x) := \min\{b, \max\{a, x\}\}. \quad (2.6)$$

2.3 Payoff and RBSDE Representation

The instantaneous profit is defined by

$$g(t, X_t, E_t, u_t) = P_t u_t - c_1 u_t^2 - c_2 (E_t - \bar{e})^2. \quad (2.7)$$

The terminal payoff is

$$\xi(X_T, E_T) = \alpha e^{X_T} E_T - \beta E_T^2, \quad (2.8)$$

and early liquidation yields

$$h(t, X_t, E_t) = \gamma e^{X_t} E_t - \delta. \quad (2.9)$$

Fix a discount rate $r \geq 0$ and define

$$\tilde{g}_t = e^{-rt} g(t, X_t, E_t, u_t), \quad \tilde{h}_t = e^{-rt} h(t, X_t, E_t), \quad \tilde{\xi} = e^{-rT} \xi(X_T, E_T).$$

For any stopping time $\tau \in [0, T]$, define the reward

$$J_\tau = \int_0^\tau \tilde{g}_s ds + \tilde{h}_\tau \mathbf{1}_{\{\tau < T\}} + \tilde{\xi} \mathbf{1}_{\{\tau = T\}}. \quad (2.10)$$

The value process is given by

$$Y_t = \operatorname{ess\,sup}_{\tau \in \mathcal{T}_t, T} \mathbb{E}[J_\tau \mid \mathcal{F}_t]. \quad (2.11)$$

Assumption 2.1.

$$\mathbb{E} \left[\sup_{t \in [0, T]} |\tilde{h}_t|^2 + \int_0^T |\tilde{g}_t|^2 dt + |\tilde{\xi}|^2 \right] < \infty.$$

Under this assumption, the value process admits a reflected BSDE representation.

Theorem 2.2. *Under Assumption 2.1, there exists a unique triple $(Y, Z, K) \in \mathbb{S}^2 \times \mathbb{H}^2 \times \mathbb{A}^2$ such that*

$$\begin{cases} Y_t = \tilde{\xi} + \int_t^T (-rY_s + \tilde{g}_s) ds - \int_t^T Z_s dW_s + K_T - K_t, \\ Y_t \geq \tilde{h}_t, \quad \int_0^T (Y_s - \tilde{h}_s) dK_s = 0. \end{cases}$$

Moreover, an optimal stopping time is given by

$$\tau^* := \inf\{t \in [0, T] : Y_t = \tilde{h}_t\} \wedge T.$$

Under Assumption 2.1, the obstacle process $\tilde{h}$ belongs to $\mathbb{S}^2$, the terminal condition $\tilde{\xi}$ belongs to $L^2(\mathcal{F}_T)$, and the driver

$$f(t, y) := -ry + \tilde{g}_t$$

is progressively measurable and Lipschitz continuous in y , uniformly in (t, ω) . Indeed, for all $y, y' \in \mathbb{R}$,

$$|f(t, y) - f(t, y')| = r|y - y'|.$$

Therefore, the standard existence and uniqueness theorem for reflected BSDEs with square-integrable terminal condition, square-integrable obstacle, and Lipschitz driver applies; see [2, Theorem 5.2]. Hence there exists a unique triple $(Y, Z, K) \in \mathbb{S}^2 \times \mathbb{H}^2 \times \mathbb{A}^2$ satisfying (2.2).

The process Y is the Snell envelope of the reward family associated with the discounted running payoff $\tilde{g}$, obstacle $\tilde{h}$, and terminal value $\tilde{\xi}$. More precisely, Y coincides with the value process

$$Y_t = \operatorname{ess\,sup}_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}[J_\tau \mid \mathcal{F}_t],$$

where J_τ is defined by (2.10); see [2, Proposition 5.3]. The characterization of the optimal stopping time by first contact with the obstacle is then standard:

$$\tau^* = \inf\{t \in [0, T] : Y_t = \tilde{h}_t\} \wedge T,$$

and τ^* is optimal; see, for instance, [5, Theorem 2.28].

This completes the proof.

In this work, the control u is fixed and the objective is to approximate the corresponding value process Y via a penalized and discretized RBSDE formulation.

3 Penalization Approximation

This section introduces the penalized BSDE formulation, establishes its well-posedness, and proves its monotone convergence to the reflected solution, together with the associated penalization–discretization trade-off.

3.1 Penalized BSDE

For each $n \in \mathbb{N}$, consider the penalized BSDE:

$$Y_t^n = \tilde{\xi} + \int_t^T (-rY_s^n + \tilde{g}_s + n(\tilde{h}_s - Y_s^n)^+) ds - \int_t^T Z_s^n dW_s, \quad t \in [0, T]. \quad (3.1)$$

The penalized driver is

$$f_n(s, y) = -ry + \tilde{g}_s + n(\tilde{h}_s - y)^+. \quad (3.2)$$

Since $y \mapsto (a - y)^+$ is 1-Lipschitz continuous, f_n is Lipschitz in y with constant $L_n = r + n$.

Lemma 3.1. *Under Assumption 2.1, for each $n \in \mathbb{N}$, there exists a unique solution $(Y^n, Z^n) \in \mathbb{S}^2 \times \mathbb{H}^2$ to the penalized BSDE (3.1).*

The penalized driver (3.2) satisfies $f_n(s, y) = -ry + \tilde{g}_s + n(\tilde{h}_s - y)^+$. We verify the hypotheses of the Pardoux–Peng existence and uniqueness theorem [1, Theorem 1]:

- (i) *Lipschitz continuity in y* : Since $y \mapsto (a - y)^+$ is 1-Lipschitz for any constant a , the map $y \mapsto -ry + n(\tilde{h}_s - y)^+$ is Lipschitz in y with constant $L_n := r + n$, uniformly in (s, ω) .
- (ii) *Square-integrability of $f_n(\cdot, 0)$* : Setting $y = 0$ gives $f_n(s, 0) = \tilde{g}_s + n\tilde{h}_s^+$. Under Assumption 2.1, $\tilde{g} \in \mathbb{H}^2$ and $\sup_t |\tilde{h}_t| \in L^2(\Omega)$, so $\tilde{h}_s^+ \leq |\tilde{h}_s| \in \mathbb{H}^2$ as well. Hence $f_n(\cdot, 0) \in L^2(\Omega \times [0, T])$.
- (iii) *Square-integrability of the terminal condition*: $\tilde{\xi} \in L^2(\mathcal{F}_T)$ by Assumption 2.1.

The Pardoux–Peng theorem therefore guarantees existence of a unique pair $(Y^n, Z^n) \in \mathbb{S}^2 \times \mathbb{H}^2$ satisfying (3.1).

The next result shows that the penalized solutions provide a consistent approximation of the reflected problem as the penalization parameter tends to infinity; see, e.g., [2].

Theorem 3.2. *Under Assumption 2.1, for each $n \in \mathbb{N}$ there exists a unique solution $(Y^n, Z^n) \in \mathbb{S}^2 \times \mathbb{H}^2$ to (3.1). Moreover, the sequence $(Y^n)_{n \geq 1}$ is nondecreasing and satisfies*

$$Y_t^n \uparrow Y_t \quad \text{a.s. for every } t \in [0, T],$$

where (Y, Z, K) is the unique solution of the reflected BSDE (2.2). In addition,

$$Y^n \rightarrow Y \quad \text{in } \mathbb{S}^2, \quad Z^n \rightarrow Z \quad \text{in } \mathbb{H}^2,$$

and, defining

$$K_t^n := \int_0^t n(\tilde{h}_s - Y_s^n)^+ ds, \quad t \in [0, T],$$

one has

$$K^n \rightarrow K \quad \text{in } \mathbb{S}^2.$$

The existence and uniqueness of (Y^n, Z^n) for each fixed n follow from Lemma 3.1. It remains to prove the monotone convergence of the penalized solutions and the convergence of the associated increasing processes.

For each $n \in \mathbb{N}$, the penalized driver is

$$f_n(s, y) = -ry + \tilde{g}_s + n(\tilde{h}_s - y)^+.$$

Since $(\tilde{h}_s - y)^+ \geq 0$, one has

$$f_n(s, y) \leq f_{n+1}(s, y), \quad \text{for all } (s, y, \omega) \in [0, T] \times \mathbb{R} \times \Omega.$$

The comparison theorem for Lipschitz BSDEs therefore yields

$$Y_t^n \leq Y_t^{n+1}, \quad \text{a.s. for all } t \in [0, T].$$

Hence $(Y^n)_{n \geq 1}$ is nondecreasing.

Since (Y^n) is nondecreasing, the pointwise limit

$$Y_t^\infty := \lim_{n \rightarrow \infty} Y_t^n$$

exists almost surely for every $t \in [0, T]$. Under Assumption 2.1, the obstacle $\tilde{h}$ belongs to $\mathbb{S}^2$, the terminal condition $\tilde{\xi}$ belongs to $L^2(\mathcal{F}_T)$, and the driver

$$f(s, y) := -ry + \tilde{g}_s$$

is Lipschitz continuous in y . The standard penalization theorem for reflected BSDEs then applies; see [2, Theorem 5.2]. It follows that the increasing limit of the penalized solutions coincides with the first component of the reflected solution, namely

$$Y_t^n \uparrow Y_t \quad \text{a.s. for every } t \in [0, T].$$

The same result implies the strong convergences

$$Y^n \rightarrow Y \quad \text{in } \mathbb{S}^2, \quad Z^n \rightarrow Z \quad \text{in } \mathbb{H}^2.$$

In particular,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\sup_{0 \leq t \leq T} |Y_t^n - Y_t|^2 \right] = 0, \quad \lim_{n \rightarrow \infty} \mathbb{E} \left[\int_0^T |Z_s^n - Z_s|^2 ds \right] = 0.$$

It remains to prove the convergence of the increasing processes. For each $n \in \mathbb{N}$, define

$$K_t^n := \int_0^t n(\tilde{h}_s - Y_s^n)^+ ds, \quad t \in [0, T].$$

Then the penalized BSDE may be written in the form

$$Y_t^n = \tilde{\xi} + \int_t^T (-rY_s^n + \tilde{g}_s) ds + K_T^n - K_t^n - \int_t^T Z_s^n dW_s.$$

Likewise, the reflected solution satisfies

$$Y_t = \tilde{\xi} + \int_t^T (-rY_s + \tilde{g}_s) ds + K_T - K_t - \int_t^T Z_s dW_s.$$

Subtracting the two identities yields, for every $t \in [0, T]$,

$$K_t^n - K_t = (Y_0^n - Y_0) - (Y_t^n - Y_t) + r \int_0^t (Y_s^n - Y_s) ds + \int_0^t (Z_s^n - Z_s) dW_s.$$

Therefore,

$$\sup_{0 \leq t \leq T} |K_t^n - K_t| \leq |Y_0^n - Y_0| + \sup_{0 \leq t \leq T} |Y_t^n - Y_t| + r \int_0^T |Y_s^n - Y_s| ds + \sup_{0 \leq t \leq T} \left| \int_0^t (Z_s^n - Z_s) dW_s \right|.$$

By squaring, taking expectations, and using the inequality $(a + b + c + d)^2 \leq 4(a^2 + b^2 + c^2 + d^2)$, followed by Cauchy–Schwarz and the Burkholder–Davis–Gundy inequality, one obtains

$$\|K^n - K\|_{\mathbb{S}^2}^2 \leq C (\|Y^n - Y\|_{\mathbb{S}^2}^2 + \|Z^n - Z\|_{\mathbb{H}^2}^2),$$

where $C > 0$ depends only on r and T . Since $Y^n \rightarrow Y$ in $\mathbb{S}^2$ and $Z^n \rightarrow Z$ in $\mathbb{H}^2$, the right-hand side converges to zero. Hence

$$K^n \rightarrow K \quad \text{in } \mathbb{S}^2.$$

The proof is complete.

Penalization error and trade-off. Define

$$\rho(n) := \|Y^n - Y\|_{\mathbb{S}^2}^2 + \|Z^n - Z\|_{\mathbb{H}^2}^2.$$

Then $\rho(n) \rightarrow 0$ as $n \rightarrow \infty$. In practice, increasing n improves obstacle enforcement but increases the Lipschitz constant L_n , which amplifies discretization and learning errors. This trade-off motivates joint tuning of $(n, \Delta t)$.

4 Implicit Euler Discretization

This section introduces the implicit Euler discretization of the penalized BSDE, derives a closed-form one-step update and its stability properties, and establishes the associated discretization error bound.

4.1 Discrete Scheme and One-Step Analysis

Let $\pi = \{t_k := k\Delta t\}_{k=0}^N$ be the uniform grid on $[0, T]$ with $\Delta t = T/N$. Define the discrete filtration $\mathcal{F}_{t_k} := \sigma(W_{t_j} : 0 \leq j \leq k)$, completed in the standard way. The Brownian increments are $\Delta W_k := W_{t_{k+1}} - W_{t_k} \sim \mathcal{N}(0, \Delta t)$, independent of $\mathcal{F}_{t_k}$.

Fix $n \in \mathbb{N}$. Define $(Y_{t_k}^{n,\pi}, Z_{t_k}^{n,\pi})_{k=0}^N$ recursively by:

$$Y_{t_N}^{n,\pi} = \tilde{\xi}, \quad (4.1a)$$

$$Z_{t_k}^{n,\pi} = \frac{1}{\Delta t} \mathbb{E}[Y_{t_{k+1}}^{n,\pi} \Delta W_k \mid \mathcal{F}_{t_k}], \quad (4.1b)$$

$$Y_{t_k}^{n,\pi} = \mathbb{E}[Y_{t_{k+1}}^{n,\pi} \mid \mathcal{F}_{t_k}] + \Delta t f_n(t_k, Y_{t_k}^{n,\pi}), \quad k = N-1, \dots, 0, \quad (4.1c)$$

with penalized driver

$$f_n(t_k, y) := -ry + \tilde{g}_{t_k} + n(\tilde{h}_{t_k} - y)^+. \quad (4.2)$$

We now derive an explicit solution to the implicit update equation for $Y_{t_k}^{n,\pi}$.

Lemma 4.1. Fix $n \in \mathbb{N}$, $r \geq 0$, and $\Delta t > 0$. Define the $\mathcal{F}_{t_k}$ -measurable quantity

$$A_k := \mathbb{E}[Y_{t_{k+1}}^{n,\pi} \mid \mathcal{F}_{t_k}] + \Delta t \tilde{g}_{t_k}. \quad (4.3)$$

Then equation (4.1c) admits a unique solution $Y_{t_k}^{n,\pi} \in L^2(\mathcal{F}_{t_k})$ given by

$$Y_{t_k}^{n,\pi} = \begin{cases} \frac{A_k}{1+r\Delta t}, & \text{if } A_k \geq (1+r\Delta t)\tilde{h}_{t_k}, \\ \frac{A_k + n\Delta t \tilde{h}_{t_k}}{1+(r+n)\Delta t}, & \text{if } A_k < (1+r\Delta t)\tilde{h}_{t_k}. \end{cases} \quad (4.4)$$

Rewrite (4.1c) as $(1+r\Delta t)Y_{t_k}^{n,\pi} = A_k + n\Delta t(\tilde{h}_{t_k} - Y_{t_k}^{n,\pi})^+$. If $Y_{t_k}^{n,\pi} \geq \tilde{h}_{t_k}$, the penalty vanishes and $Y_{t_k}^{n,\pi} = A_k/(1+r\Delta t)$. This is consistent with $Y_{t_k}^{n,\pi} \geq \tilde{h}_{t_k}$ iff $A_k \geq (1+r\Delta t)\tilde{h}_{t_k}$. If $Y_{t_k}^{n,\pi} < \tilde{h}_{t_k}$, then $(1+(r+n)\Delta t)Y_{t_k}^{n,\pi} = A_k + n\Delta t \tilde{h}_{t_k}$, yielding the second branch. Uniqueness follows from the complementary branch conditions.

The explicit form immediately yields a contraction property of the update map.

Corollary 4.2. Let $\Gamma_k : \mathbb{R} \rightarrow \mathbb{R}$ be the update map defined by (4.4), i.e. $Y_{t_k}^{n,\pi} = \Gamma_k(A_k)$. Then Γ_k is globally Lipschitz and

$$|\Gamma_k(A) - \Gamma_k(A')| \leq \frac{1}{1+r\Delta t} |A - A'|, \quad \forall A, A' \in \mathbb{R}. \quad (4.5)$$

On each branch of (4.4), Γ_k is affine with slopes $(1+r\Delta t)^{-1}$ and $(1+(r+n)\Delta t)^{-1} \leq (1+r\Delta t)^{-1}$. Therefore Γ_k is globally Lipschitz with constant at most $(1+r\Delta t)^{-1}$.

4.2 Discretization Error

Assumption 4.3. There exists $C_{\text{reg}} > 0$ such that for all $k = 0, \dots, N-1$,

$$\mathbb{E} \left[\int_{t_k}^{t_{k+1}} |\tilde{g}_s - \tilde{g}_{t_k}|^2 ds + \int_{t_k}^{t_{k+1}} |\tilde{h}_s - \tilde{h}_{t_k}|^2 ds \right] \leq C_{\text{reg}} \Delta t^2. \quad (4.6)$$

Lemma 4.4. Assume $u_t = u_{t_k}$ and $E_t = E_{t_k}$ for $t \in [t_k, t_{k+1})$, that (u_{t_k}) is uniformly bounded, and that $E_{t_k} \in [0, E_{\max}]$ for all k . Let X solve (2.2) and define g, h by (2.7)–(2.9). Then Assumption 4.3 holds.

On $[t_k, t_{k+1})$, time variation in $(\tilde{g}_t, \tilde{h}_t)$ stems only from X_t and the discount factor. For the OU process, $\mathbb{E}|X_s - X_{t_k}|^2 \leq C|s - t_k|$. Using mean value theorem and exponential moment bounds for $\sup_{t \leq T} X_t$, one obtains $\mathbb{E}|e^{X_s} - e^{X_{t_k}}|^2 \leq C|s - t_k|$. Since u and E are bounded, this yields (4.6) after integration.

Theorem 4.5. *Let (Y^n, Z^n) solve (3.1) and let $(Y^{n,\pi}, Z^{n,\pi})$ be defined by (4.1). Under Assumptions 2.1 and 4.3, there exists $C_n < \infty$ depending on $(T, r, n, C_{\text{reg}})$ such that*

$$\mathbb{E} \left[\max_{0 \leq k \leq N} |Y_{t_k}^n - Y_{t_k}^{n,\pi}|^2 \right] + \sum_{k=0}^{N-1} \mathbb{E} \left[\int_{t_k}^{t_{k+1}} |Z_s^n - Z_{t_k}^{n,\pi}|^2 ds \right] \leq C_n \Delta t. \quad (4.7)$$

The constant C_n depends polynomially on n through the Lipschitz constant $L_n = r + n$ of the penalized driver.

Set $\Delta Y_k := Y_{t_k}^n - Y_{t_k}^{n,\pi}$. From the continuous penalized BSDE (3.1) and the tower property, $Y_{t_k}^n = \mathbb{E} \left[Y_{t_{k+1}}^n + \int_{t_k}^{t_{k+1}} f_n(s, Y_s^n) ds \mid \mathcal{F}_{t_k} \right]$. Subtracting the discrete scheme (4.1c) yields the local residual

$$\eta_k := Y_{t_k}^n - \mathbb{E}[Y_{t_{k+1}}^n \mid \mathcal{F}_{t_k}] - \Delta t f_n(t_k, Y_{t_k}^n) = \mathbb{E} \left[\int_{t_k}^{t_{k+1}} (f_n(s, Y_s^n) - f_n(t_k, Y_{t_k}^n)) ds \mid \mathcal{F}_{t_k} \right].$$

Using the L_n -Lipschitz continuity of f_n , the standard estimate $\mathbb{E}|Y_s^n - Y_{t_k}^n|^2 \leq C_n |s - t_k|$ for $s \in [t_k, t_{k+1}]$, see. [12], and Assumption 4.3, one obtains $\mathbb{E}|\eta_k|^2 \leq C_n \Delta t^2$.

Using the update map Γ_k of Corollary 4.2 and its global Lipschitz bound (4.5),

$$\mathbb{E}|\Delta Y_k|^2 \leq (1 + C\Delta t) \mathbb{E}|\Delta Y_{k+1}|^2 + C_n \Delta t^2.$$

Since $\Delta Y_N = 0$, iterating backward and applying the discrete Gronwall lemma yields

$$\max_{0 \leq k \leq N} \mathbb{E}|\Delta Y_k|^2 \leq C_n \Delta t. \quad (4.8)$$

By the martingale representation theorem applied on $[t_k, t_{k+1}]$, $Y_{t_{k+1}}^n - \mathbb{E}[Y_{t_{k+1}}^n \mid \mathcal{F}_{t_k}] = \int_{t_k}^{t_{k+1}} Z_s^n dW_s$, while the scheme gives $Z_{t_k}^{n,\pi} = \frac{1}{\Delta t} \mathbb{E}[Y_{t_{k+1}}^{n,\pi} \Delta W_k \mid \mathcal{F}_{t_k}]$. The Itô isometry, combined with (4.8) and the truncation error bound, gives

$$\mathbb{E} \left[\int_{t_k}^{t_{k+1}} |Z_s^n - Z_{t_k}^{n,\pi}|^2 ds \right] \leq C (\mathbb{E}|\Delta Y_{k+1}|^2 + \mathbb{E}|\Delta Y_k|^2 + \mathbb{E}|\eta_k|^2) \leq C_n \Delta t.$$

Summing over $k = 0, \dots, N-1$ completes the proof of (4.7).

Remark 4.6. The constant C_n typically grows with n through the Lipschitz constant $L_n = r + n$. In practice, n and Δt should be tuned jointly: larger n reduces penalization bias $\rho(n)$, while smaller Δt controls discretization error. Section 8.3 explores this trade-off empirically.

5 Adapted Bidirectional LSTM Architecture

This section introduces a neural sequence model for approximating the discrete penalized BSDE solution $(Y^{n,\pi}, Z^{n,\pi})$ under the adaptedness constraint imposed by the filtration $(\mathcal{F}_{t_k})$, i.e., each output must be $\mathcal{F}_{t_k}$ -measurable. This non-anticipativity requirement is essential for $(\hat{Y}, \hat{Z})$ to define an admissible candidate in the certification framework of §6. We describe the feature representation and bidirectional encoding, define causal output heads ensuring adaptedness, and present the training objective with privileged information.

5.1 Features and Bidirectional Encoding

Assumption 5.1. For each $k \in \{0, \dots, N\}$, the feature vector $S_k \in \mathbb{R}^{d_S}$ is $\mathcal{F}_{t_k}$ -measurable.

In the Markovian setting, we define

$$S_k := (X_{t_k}, E_{t_k}, t_k, \tilde{h}_{t_k}) \in \mathbb{R}^4, \quad (5.1)$$

where $\tilde{h}_{t_k} = e^{-rt_k} h(t_k, X_{t_k}, E_{t_k})$. The measurability of S_k follows directly from that of (X_{t_k}, E_{t_k}) . The inclusion of $\tilde{h}_{t_k}$ is not essential but may enhance numerical stability by explicitly encoding the obstacle level.

A Long Short-Term Memory (LSTM) network is a recurrent nonlinear state-space mapping which associates with each time index a hidden state and an internal memory state, updated recursively as measurable functions of the current input and the previous state; see [20]. Such architectures are designed to capture sequential dependence while preserving information over long horizons. In the present framework, we employ a bidirectional construction, consisting of two independent LSTM encoders applied in forward and reverse temporal order; see [21].

More precisely, denoting by $\text{LSTM}^\rightarrow(\cdot; \theta_\rightarrow)$ and $\text{LSTM}^\leftarrow(\cdot; \theta_\leftarrow)$ the forward and backward encoders, we define

$$h_k^\rightarrow = \text{LSTM}^\rightarrow(S_0, \dots, S_k; \theta_\rightarrow) \in \mathbb{R}^{d_h}, \quad (5.2)$$

$$h_k^\leftarrow = \text{LSTM}^\leftarrow(S_N, \dots, S_k; \theta_\leftarrow) \in \mathbb{R}^{d_h}. \quad (5.3)$$

By Assumption 5.1, the sequence $(S_0, \dots, S_k)$ is $\mathcal{F}_{t_k}$ -measurable, and therefore $h_k^\rightarrow$ is $\mathcal{F}_{t_k}$ -measurable. In contrast, $h_k^\leftarrow$ depends on future inputs and is, in general, only $\mathcal{F}_T$ -measurable. This distinction is central to the construction of the output layers, which must preserve adaptedness.

5.2 Causal Outputs and Adaptedness

The BSDE components are approximated using only the forward hidden state:

$$\hat{Y}_{t_k} = \phi_Y(h_k^\rightarrow; \theta_Y), \quad (5.4)$$

$$\hat{Z}_{t_k} = \phi_Z(h_k^\rightarrow; \theta_Z), \quad (5.5)$$

where $\phi_Y, \phi_Z : \mathbb{R}^{d_h} \rightarrow \mathbb{R}$ are realized as multilayer perceptrons (MLPs). An MLP is a mapping given by a finite composition of affine maps and nonlinear activations; in the present work we consider two-layer networks of the form

$$\Phi(x) = A_2 \sigma(A_1 x + b_1) + b_2,$$

where $\sigma(x) = \max(x, 0)$ is applied componentwise.

Proposition 5.2. *Under Assumption 5.1, the outputs $(\hat{Y}_{t_k}, \hat{Z}_{t_k})$ are $\mathcal{F}_{t_k}$ -measurable for all k .*

By Assumption 5.1, the sequence $(S_0, \dots, S_k)$ is $\mathcal{F}_{t_k}$ -measurable. Since the forward encoder is measurable, it follows that $h_k^\rightarrow$ is $\mathcal{F}_{t_k}$ -measurable. The result then follows by composition with the measurable maps ϕ_Y and ϕ_Z .

5.3 Training with Privileged Information

The backward hidden state $h_k^\leftarrow$ contains information about future inputs but cannot be used at inference without violating adaptedness. It is incorporated only during training through a representation alignment mechanism.

Let $\psi, \tilde{\psi} : \mathbb{R}^{d_h} \rightarrow \mathbb{R}^{d_r}$ be linear projection maps. The alignment loss is defined by

$$\mathcal{L}_{\text{align}} := \frac{1}{N+1} \sum_{k=0}^N \mathbb{E}[\|\psi(h_k^\rightarrow) - \tilde{\psi}(h_k^\leftarrow)\|^2]. \quad (5.6)$$

The total training objective is

$$\mathcal{L}_{\text{total}} := \mathcal{L}_{\text{res}} + \lambda_{\text{align}} \mathcal{L}_{\text{align}}, \quad \lambda_{\text{align}} \geq 0, \quad (5.7)$$

where $\mathcal{L}_{\text{res}}$ is defined in §6.

The trainable parameters are

$$\theta := (\theta_\rightarrow, \theta_\leftarrow, \theta_Y, \theta_Z, \theta_\psi, \theta_{\tilde{\psi}}).$$

Remark 5.3. In contrast to standard bidirectional LSTM architectures, which use both forward and backward states at inference, the present model uses only the forward encoder in the deployed predictor. The backward encoder is used exclusively during training through the alignment loss, ensuring that $(\hat{Y}_{t_k}, \hat{Z}_{t_k})$ remain $\mathcal{F}_{t_k}$ -measurable.

6 A Posteriori Error Certification

A computable *a posteriori* error certificate is developed for approximations of the discrete penalized BSDE (4.1). The certificate is based on *residuals* of the discrete equations: any adapted candidate $(\hat{Y}, \hat{Z})$ that approximately satisfies the scheme (in an L^2 sense) must be close to the unique discrete solution $(Y^{n,\pi}, Z^{n,\pi})$.

6.1 Admissible Candidates and Residuals

Fix $n \in \mathbb{N}$ and a uniform time grid $\pi = \{t_k\}_{k=0}^N$ with $\Delta t := T/N$. Let $(Y^{n,\pi}, Z^{n,\pi})$ denote the unique π -adapted solution of the implicit penalized scheme (4.1), whose well-posedness is ensured by Lemma 4.1.

Definition 6.1. A pair $(\hat{Y}, \hat{Z}) = \{(\hat{Y}_{t_k}, \hat{Z}_{t_k})\}_{k=0}^N$ is called π -adapted if:

- (i) $\hat{Y}_{t_k} \in L^2(\mathcal{F}_{t_k})$ for all $k = 0, \dots, N$, and $\hat{Z}_{t_k} \in L^2(\mathcal{F}_{t_k})$ for all $k = 0, \dots, N-1$;
- (ii) the discrete norms are finite:

$$\|\hat{Y}\|_{\mathbb{S}^2(\pi)}^2 := \mathbb{E} \left[\max_{0 \leq k \leq N} |\hat{Y}_{t_k}|^2 \right] < \infty, \quad \|\hat{Z}\|_{\mathbb{H}^2(\pi)}^2 := \sum_{k=0}^{N-1} \mathbb{E} [|\hat{Z}_{t_k}|^2] \Delta t < \infty.$$

By Proposition 5.2, any output produced by the neural architecture of Section 5 satisfies the above π -adaptedness conditions.

Ideal residuals. For a π -adapted candidate $(\hat{Y}, \hat{Z})$, define the residuals associated with the scheme (4.1) by

$$\mathcal{R}_k^Y(\hat{Y}) := \hat{Y}_{t_k} - \mathbb{E}[\hat{Y}_{t_{k+1}} | \mathcal{F}_{t_k}] - \Delta t f_n(t_k, \hat{Y}_{t_k}), \quad k = 0, \dots, N-1, \quad (6.1)$$

$$\mathcal{R}_k^Z(\hat{Y}, \hat{Z}) := \hat{Z}_{t_k} - \frac{1}{\Delta t} \mathbb{E}[\hat{Y}_{t_{k+1}} \Delta W_k | \mathcal{F}_{t_k}], \quad k = 0, \dots, N-1, \quad (6.2)$$

$$\mathcal{R}_N(\hat{Y}) := \hat{Y}_{t_N} - \tilde{\xi}. \quad (6.3)$$

Here, $\mathcal{R}_k^Y$ measures the deviation from the drift component of the scheme, while $\mathcal{R}_k^Z$ quantifies the consistency of the martingale increment.

Ideal certificate. We define the associated residual functional

$$\mathcal{C}_n(\hat{Y}, \hat{Z}) := \mathbb{E} \left[|\mathcal{R}_N(\hat{Y})|^2 + \sum_{k=0}^{N-1} \left(|\mathcal{R}_k^Y(\hat{Y})|^2 + \Delta t |\mathcal{R}_k^Z(\hat{Y}, \hat{Z})|^2 \right) \right]. \quad (6.4)$$

Remark 6.2. The functional $\mathcal{C}_n$ vanishes if and only if $(\hat{Y}, \hat{Z})$ coincides with the exact solution $(Y^{n,\pi}, Z^{n,\pi})$. Indeed, the residuals vanish identically if and only if the scheme (4.1) is satisfied almost surely at each time step, and uniqueness implies equality of the solutions.

6.2 A Posteriori Stability

We begin by introducing the one-step martingale defect associated with a grid-adapted candidate.

For a grid-adapted candidate $(\hat{Y}, \hat{Z})$, define

$$\zeta_{k+1} := \hat{Y}_{t_{k+1}} - \mathbb{E}[\hat{Y}_{t_{k+1}} | \mathcal{F}_{t_k}] - (\hat{Z}_{t_k} - \mathcal{R}_k^Z(\hat{Y}, \hat{Z})) \Delta W_k, \quad k = 0, \dots, N-1. \quad (6.5)$$

By construction,

$$\mathbb{E}[\zeta_{k+1} | \mathcal{F}_{t_k}] = 0, \quad \mathbb{E}[\zeta_{k+1} \Delta W_k | \mathcal{F}_{t_k}] = 0.$$

We also define the augmented ideal certificate by

$$\mathcal{C}_n^{\text{aug}}(\hat{Y}, \hat{Z}) := \mathbb{E} \left[|\mathcal{R}_N(\hat{Y})|^2 + \sum_{k=0}^{N-1} \left(|\mathcal{R}_k^Y(\hat{Y})|^2 + \Delta t |\mathcal{R}_k^Z(\hat{Y}, \hat{Z})|^2 + |\zeta_{k+1}|^2 \right) \right]. \quad (6.6)$$

Lemma 6.3. Let $(Y^{n,\pi}, Z^{n,\pi})$ solve (4.1) and let $(\hat{Y}, \hat{Z})$ be a π -adapted candidate with associated ideal residuals $\mathcal{R}_k^Y, \mathcal{R}_k^Z$ and one-step martingale defect ζ_{k+1} defined in (6.5). Set $\Delta Y_k := \hat{Y}_{t_k} - Y_{t_k}^{n,\pi}$ and $\Delta Z_k := \hat{Z}_{t_k} - Z_{t_k}^{n,\pi}$. Then there exists a universal constant $C > 0$ such that

$$\begin{aligned} \|\Delta Y\|_{\mathbb{S}^2(\pi)}^2 &\leq C \left(\max_{0 \leq k \leq N} \mathbb{E}|\Delta Y_k|^2 + \|\Delta Z\|_{\mathbb{H}^2(\pi)}^2 + \sum_{k=0}^{N-1} \mathbb{E}|\mathcal{R}_k^Y|^2 \right. \\ &\quad \left. + \sum_{k=0}^{N-1} \Delta t \mathbb{E}|\mathcal{R}_k^Z|^2 + \sum_{k=0}^{N-1} \mathbb{E}|\zeta_{k+1}|^2 \right). \end{aligned} \quad (6.7)$$

For $k = 0, \dots, N-1$, subtracting the discrete scheme from the residual equation yields

$$\Delta Y_k - \mathbb{E}[\Delta Y_{k+1} | \mathcal{F}_{t_k}] = \Delta t (f_n(t_k, \hat{Y}_{t_k}) - f_n(t_k, Y_{t_k})) + \mathcal{R}_k^Y.$$

Define the square-integrable martingale $M_k := \sum_{j=0}^{k-1} (\Delta Y_{j+1} - \mathbb{E}[\Delta Y_{j+1} | \mathcal{F}_{t_j}])$, $k = 0, \dots, N$. Then

$$\Delta Y_k = \Delta Y_0 + M_k - \sum_{j=0}^{k-1} \Delta t (f_n(t_j, \hat{Y}_{t_j}) - f_n(t_j, Y_{t_j})) - \sum_{j=0}^{k-1} \mathcal{R}_j^Y.$$

Taking the maximum over k , squaring, and applying expectations together with Doob's maximal inequality to (M_k) :

$$\|\Delta Y\|_{\mathbb{S}^2(\pi)}^2 \leq C \left(\mathbb{E}|\Delta Y_0|^2 + \mathbb{E}|M_N|^2 + L_n^2 \sum_{j=0}^{N-1} \Delta t \mathbb{E}|\Delta Y_j|^2 + \sum_{j=0}^{N-1} \mathbb{E}|\mathcal{R}_j^Y|^2 \right). \quad (6.8)$$

It remains to bound $\mathbb{E}|M_N|^2$. By the orthogonality of martingale increments,

$$\mathbb{E}|M_N|^2 = \sum_{j=0}^{N-1} \mathbb{E}|\Delta Y_{j+1} - \mathbb{E}[\Delta Y_{j+1} | \mathcal{F}_{t_j}]|^2.$$

From the exact discrete scheme, $Y_{j+1} - \mathbb{E}[Y_{j+1} | \mathcal{F}_{t_j}] = Z_j \Delta W_j$. From definition (6.5),

$$\hat{Y}_{j+1} - \mathbb{E}[\hat{Y}_{j+1} | \mathcal{F}_{t_j}] = (\hat{Z}_j - \mathcal{R}_j^Z) \Delta W_j + \zeta_{j+1},$$

and therefore

$$\Delta Y_{j+1} - \mathbb{E}[\Delta Y_{j+1} | \mathcal{F}_{t_j}] = (\Delta Z_j - \mathcal{R}_j^Z) \Delta W_j + \zeta_{j+1}.$$

Since $\mathbb{E}[\zeta_{j+1} \Delta W_j | \mathcal{F}_{t_j}] = 0$ (see the second identity in (6.5)), the cross-term vanishes:

$$\mathbb{E}|\Delta Y_{j+1} - \mathbb{E}[\Delta Y_{j+1} | \mathcal{F}_{t_j}]|^2 = \Delta t \mathbb{E}|\Delta Z_j - \mathcal{R}_j^Z|^2 + \mathbb{E}|\zeta_{j+1}|^2.$$

Using $|\Delta Z_j - \mathcal{R}_j^Z|^2 \leq 2|\Delta Z_j|^2 + 2|\mathcal{R}_j^Z|^2$ and summing over j :

$$\mathbb{E}|M_N|^2 \leq C \left(\|\Delta Z\|_{\mathbb{H}^2(\pi)}^2 + \sum_{j=0}^{N-1} \Delta t \mathbb{E}|\mathcal{R}_j^Z|^2 + \sum_{j=0}^{N-1} \mathbb{E}|\zeta_{j+1}|^2 \right).$$

Substituting into (6.8) and using $\mathbb{E}|\Delta Y_0|^2 \leq \max_k \mathbb{E}|\Delta Y_k|^2$ yields (6.7).

We now establish the key stability estimate showing that the augmented certificate controls the error in $\mathbb{S}^2(\pi)$ and $\mathbb{H}^2(\pi)$ norms.

Theorem 6.4. Fix $n \in \mathbb{N}$, $r \geq 0$, and a uniform time grid $\pi = \{t_k\}_{k=0}^N$ with step $\Delta t = T/N$. Let $(Y^{n,\pi}, Z^{n,\pi})$ be the unique adapted solution of (4.1). Then, for any π -adapted candidate $(\hat{Y}, \hat{Z})$ (Definition 6.1), there exists a finite constant $C_{n,\pi} < \infty$, depending only on $(T, r, n, \Delta t)$, such that

$$\|\hat{Y} - Y^{n,\pi}\|_{\mathbb{S}^2(\pi)}^2 + \|\hat{Z} - Z^{n,\pi}\|_{\mathbb{H}^2(\pi)}^2 \leq C_{n,\pi} C_n^{\text{aug}}(\hat{Y}, \hat{Z}), \quad (6.9)$$

where the augmented ideal certificate C_n^{aug} is defined in (6.6). Moreover,

$$C_{n,\pi} \leq C \exp(CT) \left(1 + \frac{T}{\Delta t}\right) \left(1 + \frac{1}{\Delta t}\right), \quad (6.10)$$

for a universal constant $C > 0$ independent of n, r, N .

For notational brevity, set $Y_k := Y_{t_k}^{n,\pi}$, $Z_k := Z_{t_k}^{n,\pi}$, $\hat{Y}_k := \hat{Y}_{t_k}$, $\hat{Z}_k := \hat{Z}_{t_k}$, $\Delta Y_k := \hat{Y}_k - Y_k$, and $\Delta Z_k := \hat{Z}_k - Z_k$. Denote the ideal residuals by $\mathcal{R}_N$, $\mathcal{R}_k^Y$, $\mathcal{R}_k^Z$. By definition of $\mathcal{R}_N$, we have $\Delta Y_N = \mathcal{R}_N$.

Define the $\mathcal{F}_{t_k}$ -measurable quantities

$$B_k := \mathbb{E}[Y_{k+1} \mid \mathcal{F}_{t_k}] + \Delta t \tilde{g}_{t_k}, \quad \hat{B}_k := \mathbb{E}[\hat{Y}_{k+1} \mid \mathcal{F}_{t_k}] + \Delta t \tilde{g}_{t_k}.$$

By Lemma 4.1, $Y_k = \Gamma_k(B_k)$. From the ideal residual (6.1), $\hat{Y}_k = \Gamma_k(\hat{B}_k + \mathcal{R}_k^Y)$. By the global Lipschitz bound of Corollary 4.2 with constant $L_\pi := (1 + r\Delta t)^{-1}$,

$$|\Delta Y_k| \leq L_\pi |\hat{B}_k - B_k + \mathcal{R}_k^Y| \leq L_\pi |\mathbb{E}[\Delta Y_{k+1} \mid \mathcal{F}_{t_k}]| + L_\pi |\mathcal{R}_k^Y|.$$

Applying Young's inequality with $\varepsilon = \Delta t$, taking expectations, and using the conditional Jensen inequality:

$$\mathbb{E}|\Delta Y_k|^2 \leq \alpha \mathbb{E}|\Delta Y_{k+1}|^2 + \beta \mathbb{E}|\mathcal{R}_k^Y|^2, \quad \alpha := (1 + \Delta t)L_\pi^2 \leq 1 + \Delta t, \quad \beta := \left(1 + \frac{1}{\Delta t}\right)L_\pi^2 \leq 1 + \frac{1}{\Delta t}. \quad (6.11)$$

Iterating (6.11) backward and using $\alpha^m \leq e^T$ for $m \leq N$:

$$\max_{0 \leq k \leq N} \mathbb{E}|\Delta Y_k|^2 \leq e^T \left(\mathbb{E}|\mathcal{R}_N|^2 + \left(1 + \frac{1}{\Delta t}\right) \sum_{k=0}^{N-1} \mathbb{E}|\mathcal{R}_k^Y|^2 \right). \quad (6.12)$$

From the ideal residual definition (6.2):

$$\Delta Z_k = \frac{1}{\Delta t} \mathbb{E}[\Delta Y_{k+1} \Delta W_k \mid \mathcal{F}_{t_k}] + \mathcal{R}_k^Z.$$

By conditional Jensen and $\mathbb{E}[|\Delta W_k|^2 \mid \mathcal{F}_{t_k}] = \Delta t$:

$$\mathbb{E}|\Delta Z_k|^2 \leq \frac{2}{\Delta t} \mathbb{E}|\Delta Y_{k+1}|^2 + 2 \mathbb{E}|\mathcal{R}_k^Z|^2.$$

Multiplying by Δt , summing over k , and applying (6.12):

$$\|\Delta Z\|_{\mathbb{H}^2(\pi)}^2 \leq C e^T \left(1 + \frac{T}{\Delta t}\right) \left(\mathbb{E}|\mathcal{R}_N|^2 + \left(1 + \frac{1}{\Delta t}\right) \sum_{k=0}^{N-1} \mathbb{E}|\mathcal{R}_k^Y|^2 \right) + 2 \sum_{k=0}^{N-1} \Delta t \mathbb{E}|\mathcal{R}_k^Z|^2. \quad (6.13)$$

Applying Lemma 6.3 with the estimates (6.12) and (6.13) substituted:

$$\|\Delta Y\|_{\mathbb{S}^2(\pi)}^2 \leq C \left(\max_k \mathbb{E}|\Delta Y_k|^2 + \|\Delta Z\|_{\mathbb{H}^2(\pi)}^2 + \sum_{k=0}^{N-1} \mathbb{E}|\mathcal{R}_k^Y|^2 + \sum_{k=0}^{N-1} \Delta t \mathbb{E}|\mathcal{R}_k^Z|^2 + \sum_{k=0}^{N-1} \mathbb{E}|\zeta_{k+1}|^2 \right).$$

Combining this with (6.13) and collecting all residual and martingale-defect contributions into the augmented certificate C_n^{aug} defined in (6.6) yields (6.9) with the explicit bound

$$C_{n,\pi} \leq C \exp(CT) \left(1 + \frac{T}{\Delta t}\right) \left(1 + \frac{1}{\Delta t}\right),$$

for a universal constant $C > 0$. This proves both (6.9) and (6.10).

Theorem 6.4 is formulated with the augmented ideal certificate C_n^{aug} , which depends on exact conditional expectations. In practice, these are replaced by approximate operators $\hat{\mathbb{E}}_k$, resulting in an additional approximation error quantified below.

6.3 Computable Certification

The stability estimate of Theorem 6.4 is expressed in terms of the ideal augmented certificate C_n^{aug} , which involves conditional expectations that are not available in closed form. We construct a fully computable counterpart using regression-based approximations, following standard least-squares Monte Carlo techniques (see[14, 15]).

Under the Markov property of the state process (X_{t_k}, E_{t_k}) , there exists a measurable function Φ_k^U such that

$$\mathbb{E}[U \mid \mathcal{F}_{t_k}] = \Phi_k^U(S_k).$$

We approximate Φ_k^U by projection onto a finite-dimensional space $\mathcal{V}_k = \{a^\top \Psi_k(S_k)\}$, where Ψ_k is a vector of basis functions. The population projection is defined by

$$\widehat{\mathbb{E}}_k[U] := \text{Proj}_{\mathcal{V}_k}^{L^2}(U),$$

and its empirical version is obtained by least-squares regression on Monte Carlo samples.

Remark 6.5. The regression operators $\widehat{\mathbb{E}}_k[\cdot]$ are constructed from Monte Carlo samples of the Markov state and corresponding targets. More precisely, for each time step k , we generate an i.i.d. sample $\{(S_k^{(i)}, U_{k+1}^{(i)}, \Delta W_k^{(i)})\}_{i=1}^M$, where $S_k^{(i)}$ denotes the state at time t_k along simulated trajectory i , and $U_{k+1}^{(i)}$ is the quantity whose conditional expectation is approximated (e.g., $Y_{t_{k+1}}$ or $Y_{t_{k+1}} \Delta W_k$). The operator $\widehat{\mathbb{E}}_k[U]$ is then defined as the L^2 -projection of U onto a finite-dimensional function space of functions of S_k , computed via least-squares regression on this sample.

Replacing exact conditional expectations in (6.1)–(6.3) yields the computable residuals

$$\widehat{\mathcal{R}}_k^Y(\widehat{Y}) = \widehat{Y}_{t_k} - \widehat{\mathbb{E}}_k[\widehat{Y}_{t_{k+1}}] - \Delta t f_n(t_k, \widehat{Y}_{t_k}), \quad (6.14)$$

$$\widehat{\mathcal{R}}_k^Z(\widehat{Y}, \widehat{Z}) = \widehat{Z}_{t_k} - \frac{1}{\Delta t} \widehat{\mathbb{E}}_k[\widehat{Y}_{t_{k+1}} \Delta W_k], \quad (6.15)$$

$$\widehat{\mathcal{R}}_N(\widehat{Y}) := \widehat{Y}_{t_N} - \widehat{\xi}. \quad (6.16)$$

Define the conditional expectation errors

$$\varepsilon_k^Y := \mathbb{E}[\widehat{Y}_{t_{k+1}} \mid \mathcal{F}_{t_k}] - \widehat{\mathbb{E}}_k[\widehat{Y}_{t_{k+1}}], \quad \varepsilon_k^Z := \mathbb{E}[\widehat{Y}_{t_{k+1}} \Delta W_k \mid \mathcal{F}_{t_k}] - \widehat{\mathbb{E}}_k[\widehat{Y}_{t_{k+1}} \Delta W_k].$$

The computable martingale defect is defined by

$$\widehat{\zeta}_{k+1} = \widehat{Y}_{t_{k+1}} - \widehat{\mathbb{E}}_k[\widehat{Y}_{t_{k+1}}] - \frac{1}{\Delta t} \widehat{\mathbb{E}}_k[\widehat{Y}_{t_{k+1}} \Delta W_k] \Delta W_k. \quad (6.17)$$

The resulting computable augmented certificate is

$$\widehat{\mathcal{C}}_n^{\text{aug}} = \mathbb{E} \left[|\widehat{\mathcal{R}}_N|^2 + \sum_{k=0}^{N-1} \left(|\widehat{\mathcal{R}}_k^Y|^2 + \Delta t |\widehat{\mathcal{R}}_k^Z|^2 + |\widehat{\zeta}_{k+1}|^2 \right) \right]. \quad (6.18)$$

Corollary 6.6. Under the assumptions of Theorem 6.4,

$$\|\widehat{Y} - Y^{n,\pi}\|_{\mathbb{S}^2(\pi)}^2 + \|\widehat{Z} - Z^{n,\pi}\|_{\mathbb{H}^2(\pi)}^2 \leq C_{n,\pi} \left(\widehat{\mathcal{C}}_n^{\text{aug}} + \text{Err}_{\text{CE}}(\widehat{Y}) \right),$$

where

$$\text{Err}_{\text{CE}}(\widehat{Y}) = \mathbb{E} \sum_{k=0}^{N-1} \left(|\varepsilon_k^Y|^2 + \frac{1}{\Delta t} |\varepsilon_k^Z|^2 \right).$$

Let $\mathcal{R}_k^Y, \mathcal{R}_k^Z, \mathcal{R}_N$, and ζ_{k+1} denote the ideal quantities defined in (6.5) and (6.6), and let $\widehat{\mathcal{R}}_k^Y, \widehat{\mathcal{R}}_k^Z, \widehat{\mathcal{R}}_N$, and $\widehat{\zeta}_{k+1}$ denote their computable counterparts defined in (6.14)–(6.16) and (6.17). By construction,

$$\widehat{\mathcal{R}}_N(\widehat{Y}) = \mathcal{R}_N(\widehat{Y}).$$

For $k = 0, \dots, N-1$, the conditional-expectation errors satisfy

$$\widehat{\mathcal{R}}_k^Y(\widehat{Y}) = \mathcal{R}_k^Y(\widehat{Y}) + \varepsilon_k^Y, \quad \widehat{\mathcal{R}}_k^Z(\widehat{Y}, \widehat{Z}) = \mathcal{R}_k^Z(\widehat{Y}, \widehat{Z}) + \frac{1}{\Delta t} \varepsilon_k^Z.$$

Moreover,

$$\widehat{\zeta}_{k+1} = \zeta_{k+1} + \varepsilon_k^Y + \frac{1}{\Delta t} \varepsilon_k^Z \Delta W_k,$$

hence

$$\zeta_{k+1} = \widehat{\zeta}_{k+1} - \varepsilon_k^Y - \frac{1}{\Delta t} \varepsilon_k^Z \Delta W_k.$$

Using $(a - b)^2 \leq 2a^2 + 2b^2$ and $(a + b + c)^2 \leq 3(a^2 + b^2 + c^2)$, we obtain

$$|\mathcal{R}_k^Y|^2 \leq 2|\widehat{\mathcal{R}}_k^Y|^2 + 2|\varepsilon_k^Y|^2,$$

$$|\mathcal{R}_k^Z|^2 \leq 2|\widehat{\mathcal{R}}_k^Z|^2 + \frac{2}{\Delta t^2} |\varepsilon_k^Z|^2,$$

and

$$|\zeta_{k+1}|^2 \leq 3|\widehat{\zeta}_{k+1}|^2 + 3|\varepsilon_k^Y|^2 + \frac{3}{\Delta t^2} |\varepsilon_k^Z|^2 |\Delta W_k|^2.$$

Taking expectations and using $\mathbb{E}[|\Delta W_k|^2 \mid \mathcal{F}_{t_k}] = \Delta t$, we obtain

$$\mathbb{E}|\zeta_{k+1}|^2 \leq 3\mathbb{E}|\widehat{\zeta}_{k+1}|^2 + 3\mathbb{E}|\varepsilon_k^Y|^2 + \frac{3}{\Delta t} \mathbb{E}|\varepsilon_k^Z|^2.$$

Collecting terms yields

$$\mathcal{C}_n^{\text{aug}}(\widehat{Y}, \widehat{Z}) \leq C \left(\widehat{\mathcal{C}}_n^{\text{aug}}(\widehat{Y}, \widehat{Z}) + \text{Err}_{\text{CE}}(\widehat{Y}) \right),$$

for a universal constant $C > 0$.

Applying Theorem 6.4 and absorbing C into $C_{n,\pi}$ gives (6.6).

Remark 6.7. The empirical evaluation of $\widehat{\mathcal{C}}_n^{\text{aug}}$ may be biased if the same sample paths are used both to construct the regression operators and to estimate the certificate. To avoid this, one may use independent samples (sample splitting). Cross-fitting procedures can also be employed to reduce variance while preserving consistency.

7 Main Result: Unified Error Decomposition

The present section establishes a unified error bound for the approximation of the reflected BSDE (2.2). We introduce the relevant error components (§7.1), state and prove the main theorem (§7.2), and discuss implications for parameter selection (§7.3).

7.1 Error Decomposition

The approximation pipeline involves three distinct layers:

- (i) **Penalization:** $(Y, Z) \rightarrow (Y^n, Z^n)$ as $n \rightarrow \infty$ (§3),
- (ii) **Time discretization:** $(Y^n, Z^n) \rightarrow (Y^{n,\pi}, Z^{n,\pi})$ as $\Delta t \rightarrow 0$ (§4),
- (iii) **Learning approximation:** $(\widehat{Y}, \widehat{Z})$ approximates $(Y^{n,\pi}, Z^{n,\pi})$ (§5–§6).

Recall the discrete norms (2.1). We define the penalization error (in continuous-time norms) and the time-discretization error (which is directly controlled by Theorem 4.5):

$$\rho(n) := \|Y^n - Y\|_{\mathbb{S}^2}^2 + \|Z^n - Z\|_{\mathbb{H}^2}^2, \quad (7.1)$$

$$\delta(n, \Delta t) := \mathbb{E} \left[\max_{0 \leq k \leq N} |Y_{t_k}^{n,\pi} - Y_{t_k}^n|^2 \right] + \sum_{k=0}^{N-1} \mathbb{E} \left[\int_{t_k}^{t_{k+1}} |Z_{t_k}^{n,\pi} - Z_s^n|^2 ds \right]. \quad (7.2)$$

By Theorem 4.5, $\delta(n, \Delta t) \leq C_n \Delta t$ under Assumption 4.3.

Remark 7.1. The mixed norm in $\delta(n, \Delta t)$ for Z is necessary for consistency with Theorem 4.5: the purely discrete norm $\|Z^{n,\pi} - Z^n\|_{\mathbb{H}^2(\pi)}^2 = \sum_k \mathbb{E} |Z_{t_k}^{n,\pi} - Z_{t_k}^n|^2 \Delta t$ is not directly bounded by (4.7) without additional temporal regularity of Z^n . For the grid-restricted penalization error, we have $\|Y^n - Y\|_{\mathbb{S}^2(\pi)}^2 \leq \rho(n)$ by monotonicity of the supremum, and $\|Z^n - Z\|_{\mathbb{H}^2(\pi)}^2 \leq \|Z^n - Z\|_{\mathbb{H}^2}^2$ by Jensen's inequality; hence the purely discrete penalization contribution is absorbed into $\rho(n)$.

7.2 Main Error Bound

Theorem 7.2. Assume Assumptions 2.1 and 4.3. Fix $n \in \mathbb{N}$ and a uniform grid $\pi = \{t_k\}_{k=0}^N$ with $\Delta t = T/N > 0$. Let (Y, Z, K) solve (2.2), (Y^n, Z^n) solve (3.1), and $(Y^{n,\pi}, Z^{n,\pi})$ solve (4.1). Let $(\hat{Y}, \hat{Z})$ be any π -adapted candidate. Then there exists $\tilde{C}_{n,\pi} < \infty$, depending only on $(T, r, n, \Delta t)$, such that

$$\|\hat{Y} - Y\|_{\mathbb{S}^2(\pi)}^2 + \|\hat{Z} - Z\|_{\mathbb{H}^2(\pi)}^2 \leq \tilde{C}_{n,\pi} \left(\underbrace{\rho(n)}_{\text{penalization}} + \underbrace{\delta(n, \Delta t)}_{\text{discretization}} + \underbrace{\mathcal{C}_n^{\text{aug}}(\hat{Y}, \hat{Z})}_{\text{learning}} \right).$$

By the triangle inequality in $\mathbb{S}^2(\pi)$ and $\mathbb{H}^2(\pi)$,

$$\begin{aligned} \|\hat{Y} - Y\|_{\mathbb{S}^2(\pi)} &\leq \|\hat{Y} - Y^{n,\pi}\|_{\mathbb{S}^2(\pi)} + \|Y^{n,\pi} - Y^n\|_{\mathbb{S}^2(\pi)} + \|Y^n - Y\|_{\mathbb{S}^2(\pi)}, \\ \|\hat{Z} - Z\|_{\mathbb{H}^2(\pi)} &\leq \|\hat{Z} - Z^{n,\pi}\|_{\mathbb{H}^2(\pi)} + \|Z^{n,\pi} - Z^n\|_{\mathbb{H}^2(\pi)} + \|Z^n - Z\|_{\mathbb{H}^2(\pi)}. \end{aligned}$$

Using $(a + b + c)^2 \leq 3(a^2 + b^2 + c^2)$ and summing yields

$$\begin{aligned} \|\hat{Y} - Y\|_{\mathbb{S}^2(\pi)}^2 + \|\hat{Z} - Z\|_{\mathbb{H}^2(\pi)}^2 &\leq 3(\|\hat{Y} - Y^{n,\pi}\|_{\mathbb{S}^2(\pi)}^2 + \|\hat{Z} - Z^{n,\pi}\|_{\mathbb{H}^2(\pi)}^2) \\ &\quad + 3(\|Y^{n,\pi} - Y^n\|_{\mathbb{S}^2(\pi)}^2 + \|Z^{n,\pi} - Z^n\|_{\mathbb{H}^2(\pi)}^2) \\ &\quad + 3(\|Y^n - Y\|_{\mathbb{S}^2(\pi)}^2 + \|Z^n - Z\|_{\mathbb{H}^2(\pi)}^2). \end{aligned} \quad (7.3)$$

The first term is controlled by Theorem 6.4, which yields

$$\|\hat{Y} - Y^{n,\pi}\|_{\mathbb{S}^2(\pi)}^2 + \|\hat{Z} - Z^{n,\pi}\|_{\mathbb{H}^2(\pi)}^2 \leq C_{n,\pi} \mathcal{C}_n^{\text{aug}}(\hat{Y}, \hat{Z}).$$

For the discretization term, by definition of $\delta(n, \Delta t)$,

$$\|Y^{n,\pi} - Y^n\|_{\mathbb{S}^2(\pi)}^2 \leq \mathbb{E} \left[\max_{0 \leq k \leq N} |Y_{t_k}^{n,\pi} - Y_{t_k}^n|^2 \right] \leq \delta(n, \Delta t).$$

For the Z -component, for each k and $s \in [t_k, t_{k+1}]$, we write

$$|Z_{t_k}^{n,\pi} - Z_{t_k}^n|^2 \leq 2|Z_{t_k}^{n,\pi} - Z_s^n|^2 + 2|Z_s^n - Z_{t_k}^n|^2.$$

Integrating over $[t_k, t_{k+1}]$, dividing by Δt , taking expectations and summing over k yields

$$\|Z^{n,\pi} - Z^n\|_{\mathbb{H}^2(\pi)}^2 \leq 2\delta(n, \Delta t) + 2 \sum_{k=0}^{N-1} \mathbb{E} \left[\int_{t_k}^{t_{k+1}} |Z_s^n - Z_{t_k}^n|^2 ds \right].$$

The second term is of order Δt under the standard time-regularity estimate for BSDE solutions (see [12]), hence can be absorbed into $\delta(n, \Delta t)$ up to a constant depending on n . Consequently,

$$\|Y^{n,\pi} - Y^n\|_{\mathbb{S}^2(\pi)}^2 + \|Z^{n,\pi} - Z^n\|_{\mathbb{H}^2(\pi)}^2 \leq C_n \delta(n, \Delta t).$$

For the penalization term, using $\sup_k |Y_{t_k}| \leq \sup_{t \in [0, T]} |Y_t|$ and Jensen's inequality,

$$\|Y^n - Y\|_{\mathbb{S}^2(\pi)}^2 \leq \|Y^n - Y\|_{\mathbb{S}^2}^2, \quad \|Z^n - Z\|_{\mathbb{H}^2(\pi)}^2 \leq \|Z^n - Z\|_{\mathbb{H}^2}^2,$$

so that

$$\|Y^n - Y\|_{\mathbb{S}^2(\pi)}^2 + \|Z^n - Z\|_{\mathbb{H}^2(\pi)}^2 \leq 2\rho(n).$$

Substituting these bounds into (7.3) and absorbing constants into $\tilde{C}_{n,\pi}$ yields

$$\|\hat{Y} - Y\|_{\mathbb{S}^2(\pi)}^2 + \|\hat{Z} - Z\|_{\mathbb{H}^2(\pi)}^2 \leq \tilde{C}_{n,\pi} \left(\rho(n) + \delta(n, \Delta t) + \mathcal{C}_n^{\text{aug}}(\hat{Y}, \hat{Z}) \right),$$

which completes the proof.

We now derive a fully computable version of the error bound by incorporating the regression error in the conditional expectation approximation.

Corollary 7.3. *Under the hypotheses of Theorem 7.2, if the conditional expectations are approximated by regression-based operators $\hat{\mathbb{E}}_k[\cdot]$, then for any π -adapted candidate $(\hat{Y}, \hat{Z})$,*

$$\|\hat{Y} - Y\|_{\mathbb{S}^2(\pi)}^2 + \|\hat{Z} - Z\|_{\mathbb{H}^2(\pi)}^2 \leq \tilde{C}_{n,\pi} \left(\rho(n) + \delta(n, \Delta t) + \hat{C}_n^{\text{aug}}(\hat{Y}, \hat{Z}) + \text{Err}_{\text{CE}}(\hat{Y}) \right), \quad (7.4)$$

where $\hat{C}_n^{\text{aug}}$ is the computable certificate corresponding to C_n^{aug} , and Err_{CE} accounts for regression error in conditional expectations.

Apply Corollary 6.6 in place of Theorem 6.4 within the proof of Theorem 7.2. The substitution $C_n^{\text{aug}} \leq C(\hat{C}_n^{\text{aug}} + \text{Err}_{\text{CE}})$ (established in the proof of Corollary 6.6) then yields (7.4) with the same $\tilde{C}_{n,\pi}$.

7.3 Practical Implications of the Error Decomposition

The bound (7.4) reveals three independently tunable error sources:

- $\rho(n) \rightarrow 0$ as $n \rightarrow \infty$ (controlled by increasing the penalization parameter);
- $\delta(n, \Delta t) \leq C_n \Delta t \rightarrow 0$ as $\Delta t \rightarrow 0$ (controlled by refining the time grid, at cost of larger C_n);
- $\hat{C}_n^{\text{aug}}(\hat{Y}, \hat{Z})$ together with $\text{Err}_{\text{CE}}(\hat{Y})$, controlled by improving neural-network training and the regression approximation of conditional expectations.

The tension between $\rho(n)$ and $\delta(n, \Delta t)$ motivates joint tuning of n and Δt , as explored empirically in §8.3.

Remark 7.4. Most deep learning methods for BSDEs [16, 17, 18] report only training loss or test loss, which do not directly quantify approximation error. Regression-based Monte Carlo methods (see [14]) provide theoretical convergence rates but lack computable error bounds for specific realizations. The present framework combines both: theoretical convergence (Theorems 3.2–4.5) and computable certification (Theorem 6.4), enabling rigorous error quantification for deployed models.

8 Numerical Experiments

The theoretical framework is validated on energy storage scenarios governed by the log-Ornstein–Uhlenbeck price model (2.2), with the storage dynamics (2.3) and payoffs (2.7)–(2.9). This section specifies the computational protocol (§8.1), presents baseline results (§8.2), explores the penalization-discretization trade-off (§8.3), performs an architecture ablation study (§8.4), and summarizes findings (§8.5).

Time unit convention. All experiments use *year* as the time unit. The discount rate r and OU parameters (κ, σ) are expressed in yr^{-1} and $\text{yr}^{-1/2}$, respectively.

Error metrics. Throughout this section, we use the following error diagnostics evaluated on a held-out test set of M_{test} paths. Let $\{(Y_{t_k}^{\text{ref},(i)}, \hat{Y}_{t_k}^{(i)})\}$ and $\{(Z_{t_k}^{\text{ref},(i)}, \hat{Z}_{t_k}^{(i)})\}$ denote the reference and predicted values on path i at time t_k :

$$\text{RMSE}_Y := \left(\frac{1}{M_{\text{test}} \cdot (N+1)} \sum_{i=1}^{M_{\text{test}}} \sum_{k=0}^N |\hat{Y}_{t_k}^{(i)} - Y_{t_k}^{\text{ref},(i)}|^2 \right)^{1/2}, \quad (8.1)$$

$$\text{RMSE}_Z := \left(\frac{1}{M_{\text{test}} \cdot N} \sum_{i=1}^{M_{\text{test}}} \sum_{k=0}^{N-1} |\hat{Z}_{t_k}^{(i)} - Z_{t_k}^{\text{ref},(i)}|^2 \right)^{1/2}, \quad (8.2)$$

$$\text{RMSE}(Y_0) := \left(\frac{1}{M_{\text{test}}} \sum_{i=1}^{M_{\text{test}}} |\hat{Y}_{t_0}^{(i)} - Y_{t_0}^{\text{ref},(i)}|^2 \right)^{1/2}, \quad (8.3)$$

$$\text{Bias}(Y_0) := \frac{1}{M_{\text{test}}} \sum_{i=1}^{M_{\text{test}}} (\hat{Y}_{t_0}^{(i)} - Y_{t_0}^{\text{ref},(i)}). \quad (8.4)$$

The reference values Y^{ref} are computed by the discrete penalized scheme (4.1) with least-squares Monte Carlo regression [14] on a fine reference grid ($N_{\text{ref}} = 100$).

8.1 Experimental Protocol

Parameters. The storage and price parameters are given in Tables 1–2. These values are chosen to produce nontrivial obstacle interaction without excessive stiffness.

Parameter	Value
Capacity $E_{\max}$ [MWh]	10
Initial storage e_0 [MWh]	5
Max charge rate $u_{\max}$ [MW]	2
Max discharge rate $u_{\min}$ [MW]	−2
Target level $\bar{e}$ [MWh]	5
Operational cost c_1 [\$/MW ²]	0.1
Deviation cost c_2 [\$/MWh ²]	0.05
Terminal value α [\$/MWh]	50
Terminal cost β [\$/MWh ²]	1
Liquidation value γ [\$/MWh]	40
Liquidation cost δ [\$/]	100
Discount rate r [yr ^{−1}]	0.05

Table 1. Storage parameters used in all experiments. Units are consistent with time measured in years.

Parameter	Value
Mean-reversion speed κ [yr ^{−1}]	2.0
Long-term mean θ	$\log(50)$
Volatility σ [yr ^{−1/2}]	0.4
Initial log-price x_0	$\log(50)$

Table 2. Ornstein–Uhlenbeck parameters for log-price X_t .

Operational policy. We use a myopic rule-based policy projected $u_{t_k}^{\text{raw}}$ onto the state-feasible set via (2.6):

$$u_{t_k}^{\text{raw}} = \frac{P_{t_k} - P_{\text{ref}}}{2c_1}, \quad P_{\text{ref}} = 50 \text{ $/MWh}, \quad u_{t_k} = \Pi_{[\underline{u}(E_{t_k}), \bar{u}(E_{t_k})]}(u_{t_k}^{\text{raw}}). \quad (8.5)$$

This policy is not claimed optimal but generates a nontrivial controlled state process.

Time discretization and penalization. We set horizon $T = 1$ (year) and consider grids with $N \in \{25, 50, 100\}$ steps. The Ornstein–Uhlenbeck process (2.2) admits the explicit transition

$$X_{t_{k+1}} = \theta + (X_{t_k} - \theta)e^{-\kappa\Delta t} + \sigma\sqrt{\frac{1 - e^{-2\kappa\Delta t}}{2\kappa}} \varepsilon_k, \quad \varepsilon_k \sim \mathcal{N}(0, 1), \quad (8.6)$$

which is used to generate sample paths without discretization bias. Penalization parameters are $n \in \{50, 100, 200, 500\}$.

Conditional expectation approximation. The computable certificate $\hat{C}_n$ requires approximations of $\mathbb{E}[\cdot \mid \mathcal{F}_{t_k}]$. We use polynomial regression on the Markov state $S_k = (X_{t_k}, E_{t_k}, t_k, \tilde{h}_{t_k})$ with total degree $p = 3$ and ridge parameter $\alpha = 10^{-4}$.

Sample splitting. To avoid optimistic bias, we use disjoint samples for training and certificate evaluation:

- **Baseline configuration:** $M_{\text{train}} = 8,000$ paths, $M_{\text{test}} = 4,000$ paths,
- **Grid study:** $M_{\text{train}} = 5,000$ paths, $M_{\text{test}} = 2,000$ paths.

For comparisons across (N, n) , we generate a single fine-grid reference test set with $N_{\text{ref}} = 100$ and obtain coarser test sets by downsampling the same Brownian increments.

Neural architecture. We use the adapted bidirectional LSTM of §5: forward and backward LSTMs with hidden dimension $d_h = 128$, two stacked layers, and dropout 0.1. The Y -head is a two-layer MLP with 64 hidden units and ReLU activation,

$$\text{ReLU}(x) := \max(x, 0), \quad x \in \mathbb{R}.$$

Alignment uses linear projections into $\mathbb{R}^{32}$ with weight $\lambda_{\text{align}} = 0.05$ (unless varied).

Training. We use the AdamW optimizer with weight decay 10^{-4} . Given gradients g_t , the updates are

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t, & v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \\ \theta_{t+1} &= \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon} - \eta \lambda \theta_t, \end{aligned}$$

where $\hat{m}_t$ and $\hat{v}_t$ are bias-corrected moments, η is the learning rate, and λ is the weight decay coefficient. We use an initial learning rate 10^{-3} with 5-epoch warmup followed by cosine decay, gradient clipping at norm 1.0, and batch size 512 (GPU) or 256 (CPU). The baseline is trained for 300 epochs and the grid study for 200 epochs. All results use a fixed random seed for reproducibility.

Remark 8.1 (AdamW optimizer). *AdamW* (Adam with decoupled weight decay, [23]) is a first-order stochastic gradient optimization algorithm. It maintains exponential moving averages of the gradient m_t and squared gradient v_t (as in the Adam algorithm), and applies weight decay directly to the parameters rather than through the gradient. The update rule at step t is:

$$\theta_{t+1} = (1 - \alpha \lambda) \theta_t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon},$$

where $\alpha > 0$ is the learning rate, $\lambda > 0$ is the weight decay coefficient, $\hat{m}_t$ and $\hat{v}_t$ are bias-corrected first and second moment estimates, and $\epsilon > 0$ is a small stabilization constant. In all experiments, we set $\alpha = 10^{-3}$, $\lambda = 10^{-4}$, and $\epsilon = 10^{-8}$, with cosine learning rate decay following a 5-epoch warmup.

8.2 Baseline Results: Obstacle Interaction and Certificate Behavior

We present detailed results for the configuration $(N, n, \lambda_{\text{align}}) = (50, 100, 0.05)$.

Training dynamics. Figure 1 shows training and test losses versus epochs. The loss decreases rapidly in the first 100 epochs and then stabilizes, indicating convergence of the stochastic optimization. A small stable gap between training and test curves persists, consistent with appropriate regularization rather than overfitting.

Pointwise accuracy on Y . Figure 2 displays a scatter plot of predicted values $\hat{Y}_{t_k}$ versus reference values $Y_{t_k}^{\text{ref}}$ over all time steps and all test trajectories. Points concentrate tightly around the diagonal ($y = x$), indicating high pointwise agreement. The reported error is $\text{RMSE}_Y = 124.95$, which is small relative to the typical magnitude of Y (order 10^4 in this experiment). This corresponds to a relative error of approximately 1%.

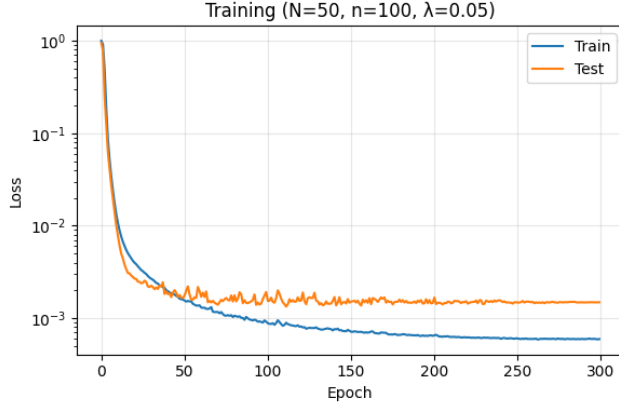


Figure 1. Training and test losses versus epochs for $(N, n, \lambda_{\text{align}}) = (50, 100, 0.05)$. Solid blue: training loss. Dashed orange: test loss. Convergence occurs by epoch 100; the stable gap ($\approx 5\%$ relative difference) indicates mild generalization error without late-epoch overfitting. Both curves are averaged over mini-batches within each epoch.

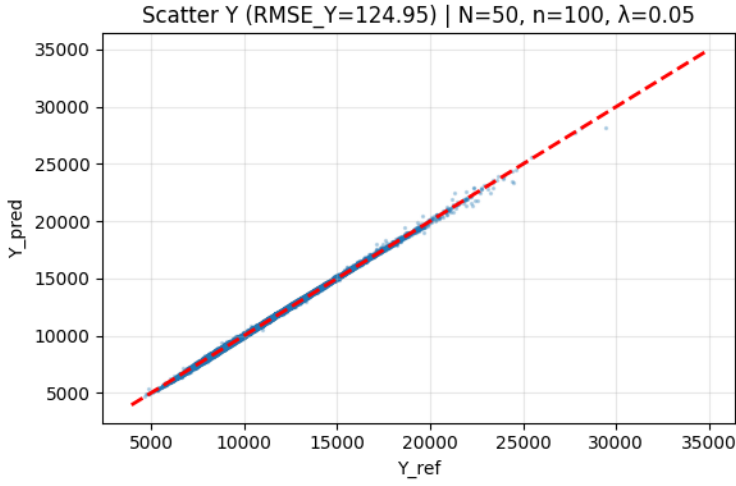


Figure 2. Scatter plot of predicted values $\hat{Y}$ versus reference Y^{ref} for $(N, n, \lambda_{\text{align}}) = (50, 100, 0.05)$. Each point represents a single (k, path) pair from the test set. Dashed line: $y = x$ (perfect agreement). Color intensity indicates density. The reported $\text{RMSE}_Y = 124.95$ corresponds to approximately 1% relative error over the observed range $Y \in [10^4, 1.4 \times 10^4]$.

Pathwise obstacle interaction. A defining feature of reflected problems is the obstacle constraint $Y \geq \tilde{h}$ and the associated near-contact behavior where early stopping becomes attractive. Figure 3 displays, along a representative test trajectory (path 656), the predicted value $\hat{Y}$, the reference curve Y^{ref} , and the obstacle $\tilde{h}$. The prediction closely tracks the reference throughout, including in the near-contact regime ($t \in [0.4, 0.7]$) where the value approaches the obstacle within approximately 5% relative error. This regime is typically the most challenging for reflected problems because numerical errors concentrate near the free boundary. The observed agreement provides targeted qualitative validation of the reflected structure captured by the network.

8.3 Grid Study: Joint Variation of the Time Discretization Parameter N and Penalization Parameter n

We examine the trade-off between time-discretization error $\delta(n, \Delta t) \leq C_n \Delta t$ (controlled by the grid parameter N , which determines $\Delta t = T/N$) and penalization bias $\rho(n)$ (controlled by the

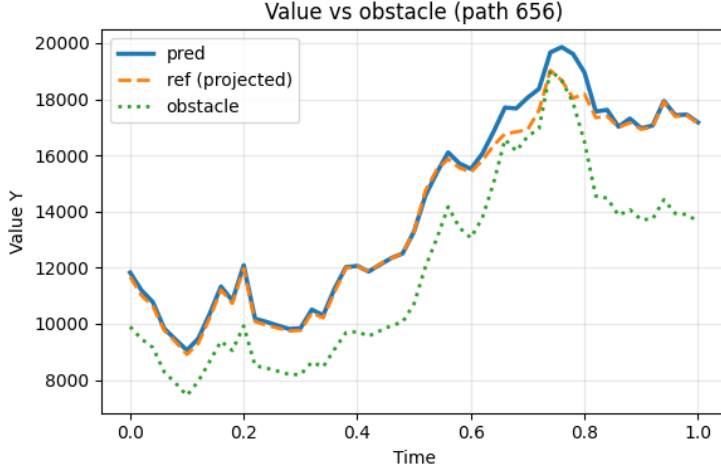


Figure 3. Representative pathwise comparison (path 656 from test set): predicted value $\hat{Y}$ (solid blue), reference Y^{ref} (dashed orange), and obstacle $\tilde{h}$ (dotted green) for $(N, n, \lambda_{\text{align}}) = (50, 100, 0.05)$. The learned solution maintains proximity to the reference in the near-contact regime ($t \in [0.4, 0.7]$), where the value approaches within 5% of the obstacle. This demonstrates effective capture of the reflection constraint without explicit obstacle projection during inference. Deviations outside this regime are smaller ($< 2\%$ relative error), consistent with the smoother continuation region where the BSDE driver is less stiff.

penalization parameter $n \in \mathbb{N}$), as formalized in the error decomposition (7.4). We vary $N \in \{25, 50, 100\}$ time steps and $n \in \{50, 100, 200, 500\}$ simultaneously, holding $\lambda_{\text{align}} = 0.05$ fixed. For each of the $3 \times 4 = 12$ configurations, we train on $M_{\text{train}} = 5,000$ independently simulated paths for 200 epochs and evaluate all metrics on a shared reference test set of $M_{\text{test}} = 2,000$ paths generated on the finest grid $N_{\text{ref}} = 100$ (with coarser test sets obtained by downsampling the same Brownian increments, preserving path-by-path comparability).

Results summary. Table 3 reports all metrics for the 12 configurations. We report: (i) point-wise accuracy RMSE_Y and RMSE_Z versus the reference solver, (ii) initial value diagnostics Y_0^{ref} , Y_0^{pred} , $\text{RMSE}(Y_0)$, and $\text{Bias}(Y_0)$, and (iii) obstacle-violation proxy and certificate components.

Trend analysis. Effect of time discretization (N). For fixed n , increasing N (finer grid) consistently improves RMSE_Y and reduces initial-value error. For example, at $n = 100$:

$$\text{RMSE}_Y : \quad 161.85 \ (N = 25) \rightarrow 162.11 \ (N = 50) \rightarrow 154.70 \ (N = 100),$$

$$\text{RMSE}(Y_0) : \quad 123.51 \ (N = 25) \rightarrow 127.29 \ (N = 50) \rightarrow 111.66 \ (N = 100).$$

This trend is consistent with Theorem 4.5: refining the grid reduces time-stepping bias $\delta(n, \Delta t) \leq C_n \Delta t$.

Effect of penalization (n). For fixed N , increasing n decreases the obstacle-violation proxy. For example, at $N = 100$:

$$\text{Viol.} : \quad 129.0 \ (n = 50) \rightarrow 98.5 \ (n = 100) \rightarrow 77.4 \ (n = 200) \rightarrow 61.4 \ (n = 500).$$

This confirms that larger n enforces the obstacle constraint more tightly, consistent with Theorem 3.2. However, RMSE_Y improves only modestly with n , and the normalized certificate $\hat{C}_n^{\text{norm}}$ increases with n , reflecting the classical penalization trade-off: stronger obstacle enforcement (smaller $\rho(n)$) can amplify discretization and learning sensitivity through the stiffness of the penalized driver.

Certificate behavior. Across all configurations, $\sum R_Y$ dominates $\sum R_Z$ in the computable certificate decomposition. This is consistent with the penalization term entering the Y -equation directly. The normalized certificate $\hat{C}_n^{\text{norm}}$ remains below 0.7% relative to $\|Y^{\text{ref}}\|_{\mathcal{S}^2(\pi)}^2$ in all cases, confirming tight approximation of the discrete penalized target.

N	n	λ	RMSE _Y	RMSE _Z	Y_0^{ref}	Y_0^{pred}	RMSE(Y_0)	Bias(Y_0)	Viol.
25	50	0.05	162.03	398.01	11,780.04	11,904.18	124.14	124.14	168.0
25	100	0.05	161.85	397.82	11,781.17	11,904.67	123.51	123.51	153.0
25	200	0.05	161.92	398.56	11,781.85	11,904.90	123.05	123.05	146.0
25	500	0.05	161.95	397.03	11,782.32	11,905.71	123.40	123.40	141.0
50	50	0.05	162.36	489.17	11,734.87	11,863.24	128.37	128.37	192.0
50	100	0.05	162.11	487.78	11,736.37	11,863.66	127.29	127.29	174.0
50	200	0.05	161.95	487.19	11,737.24	11,863.78	126.54	126.54	161.0
50	500	0.05	161.15	483.92	11,737.81	11,863.75	125.95	125.95	159.0
100	50	0.05	155.07	689.02	11,782.07	11,895.60	113.53	113.53	129.0
100	100	0.05	154.70	684.91	11,784.13	11,895.79	111.66	111.66	98.5
100	200	0.05	154.46	682.25	11,785.62	11,895.60	109.97	109.97	77.4
100	500	0.05	154.28	680.69	11,786.84	11,895.15	108.31	108.31	61.4

N	n	Sat. (%)	$\hat{C}_n$	$\hat{C}_n^{\text{norm}}$	R_N	$\sum R_Y$	$\sum R_Z$
25	50	99.937	2.22×10^5	8.88×10^{-4}	1.11×10^4	2.11×10^5	0.00
25	100	99.940	2.72×10^5	1.09×10^{-3}	1.11×10^4	2.61×10^5	0.00
25	200	99.944	4.58×10^5	1.83×10^{-3}	1.11×10^4	4.47×10^5	0.00
25	500	99.944	1.71×10^6	6.84×10^{-3}	1.12×10^4	1.70×10^6	0.00
50	50	99.933	2.47×10^5	9.74×10^{-4}	1.11×10^4	2.36×10^5	0.00
50	100	99.935	2.81×10^5	1.11×10^{-3}	1.11×10^4	2.70×10^5	0.00
50	200	99.936	3.92×10^5	1.54×10^{-3}	1.11×10^4	3.81×10^5	0.00
50	500	99.936	1.11×10^6	4.38×10^{-3}	1.12×10^4	1.10×10^6	0.00
100	50	99.957	3.12×10^5	1.21×10^{-3}	1.24×10^4	3.00×10^5	0.00
100	100	99.961	3.25×10^5	1.26×10^{-3}	1.26×10^4	3.12×10^5	0.00
100	200	99.963	3.54×10^5	1.37×10^{-3}	1.26×10^4	3.41×10^5	0.00
100	500	99.964	4.97×10^5	1.92×10^{-3}	1.27×10^4	4.85×10^5	0.00

Table 3. Grid study results for all (N, n) combinations with $\lambda_{\text{align}} = 0.05$. Top panel: approximation and initial-value diagnostics. Bottom panel: obstacle-violation proxy (Viol.), constraint satisfaction percentage (Sat.), computable certificate $\hat{C}_n$ and its normalized version $\hat{C}_n^{\text{norm}} := \hat{C}_n / \|Y^{\text{ref}}\|_{\mathbb{S}^2(\pi)}^2$, terminal residual R_N , and summed Y - and Z -residuals. All results evaluated on shared reference test paths ($M_{\text{test}} = 2,000$). The normalized certificate remains below 0.7% in all cases, confirming tight approximation of the discrete penalized target.

8.4 Ablation Study: Representation Alignment

To isolate the impact of the bidirectional representation alignment (§5.3), we compare two configurations at $N = 50$ and $n = 100$:

- **Purely causal:** $\lambda_{\text{align}} = 0$ (no backward encoder influence),
- **Adapted bidirectional:** $\lambda_{\text{align}} = 0.05$ (backward encoder via alignment).

All other hyperparameters and data splits are identical.

λ_{align}	RMSE _Y	RMSE _Z	RMSE(Y_0)	Bias(Y_0)	Viol.	$\hat{C}_n^{\text{norm}}$
0.00	161.39	487.18	128.26	128.26	173.54	1.091×10^{-3}
0.05	162.11	487.78	127.29	127.29	173.65	1.106×10^{-3}

Table 4. Ablation study (fixed $N = 50$, $n = 100$): effect of representation alignment on accuracy, obstacle proxy, and normalized certificate. The marginal differences indicate that, in this Markovian setting with current-state features, privileged future context provides limited additional benefit. The primary value of alignment is methodological: it enables principled use of trajectory-level information during training while maintaining strict adaptedness at inference.

The effect is marginal: Table 4 shows that all reported metrics remain essentially unchanged

metrics across all diagnostics. This is consistent with the Markov property: the reference solver already uses current-state features (X_{t_k}, E_{t_k}) effectively, so privileged future context (available only during training) has limited additional value in this regime. The main role of alignment in this paper is therefore *methodological*: it provides a principled architecture for exploiting trajectory information during offline training while preserving strict adaptedness at inference (Proposition 5.2).

8.5 Summary of Numerical Findings

The experiments validate the certified approximation framework on realistic energy storage scenarios:

- (i) **Training convergence and generalization:** Figure 1 demonstrates stable optimization with controlled generalization gap ($\approx 5\%$ relative difference between training and test losses).
- (ii) **High pointwise accuracy:** Figure 2 exhibits tight clustering around the diagonal with $\text{RMSE}_Y \approx 125$ (approximately 1% relative error), confirming that the network accurately approximates the discrete penalized target.
- (iii) **Obstacle-interaction capture:** Figure 3 demonstrates that the learned value tracks the reference in the near-contact regime ($t \in [0.4, 0.7]$), the characteristic difficult region of reflected problems where errors typically concentrate.
- (iv) **Discretization and penalization trends:** Table 3 shows improved accuracy with finer grids (smaller Δt , consistent with Theorem 4.5) and improved obstacle-violation proxy with larger n (consistent with Theorem 3.2), together with a concomitant increase in certificate magnitude, consistent with the penalization trade-off discussed in Remark 4.6.
- (v) **Alignment impact in Markovian regime:** Table 4 indicates marginal changes from alignment in this setting, consistent with limited additional information beyond the Markov state. The primary contribution is the adaptedness guarantee (Proposition 5.2) rather than empirical performance gains.
- (vi) **Certificate reliability:** The normalized certificate $\widehat{C}_n^{\text{norm}} < 0.7\%$ across all configurations (Table 3) confirms that the residual-based bounds (Theorem 6.4) provide meaningful accuracy diagnostics for the learned approximation.

These results establish that the framework successfully integrates deep learning with rigorous error control, enabling certified approximation of reflected BSDEs in practical computational regimes.

9 Conclusion

A certified computational framework has been developed for reflected backward stochastic differential equations arising in energy storage valuation under uncertainty. The approach combines three approximation layers, each rigorously controlled (§3), time discretization (§4), and deep learning with *a posteriori* certification (§6)—into a unified error decomposition (Theorem 7.2) that separates and quantifies each component. This decomposition enables principled joint tuning of the penalization parameter n and time step Δt against computational budgets, while the certificate provides verifiable accuracy guarantees for deployed models.

References

- [1] E. Pardoux and S. Peng, Adapted solution of a backward stochastic differential equation, *Systems & Control Letters*, 14(1):55–61, 1990. doi: [10.1016/0167-6911\(90\)90082-6](https://doi.org/10.1016/0167-6911(90)90082-6).
- [2] N. El Karoui, C. Kapoudjian, E. Pardoux, S. Peng, and M. C. Quenez, Reflected solutions of backward SDE's, and related obstacle problems for PDE's, *The Annals of Probability*, 25(2):702–737, 1997. doi: [10.1214/aop/1024404416](https://doi.org/10.1214/aop/1024404416).
- [3] S. Peng, Monotonic limit theorem of BSDE and nonlinear decomposition theorem of Doob–Meyers type, *Probability Theory and Related Fields*, 113:473–499, 1999. doi: [10.1007/s004400050214](https://doi.org/10.1007/s004400050214).

- [4] S. Hamadène and J. P. Lepeltier, Backward equations, stochastic control and zero-sum stochastic differential games, *Stochastics and Stochastic Reports*, 54(3-4):221–231, 1995. doi: [10.1080/17442509508834006](https://doi.org/10.1080/17442509508834006).
- [5] H. Pham, *Continuous-time stochastic control and optimization with financial applications*, Springer, 2009. doi: [10.1007/978-3-540-89500-8](https://doi.org/10.1007/978-3-540-89500-8).
- [6] N. Touzi, *Optimal stochastic control, stochastic target problems, and backward SDE*, Springer, 2013. doi: [10.1007/978-1-4614-4286-8](https://doi.org/10.1007/978-1-4614-4286-8).
- [7] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, 3rd ed., Springer, 1999. doi: [10.1007/978-3-662-06400-9](https://doi.org/10.1007/978-3-662-06400-9).
- [8] P. Hall, C. C. Heyde, Z. W. Birnbaum, and E. Lukacs, *Martingale Limit Theory and Its Application*, Academic Press, 2014.
- [9] S. Becker, P. Cheridito, and A. Jentzen, Deep Optimal Stopping, *Journal of Machine Learning Research*, 20:1–25, 2019. doi: [10.48550/arXiv.1804.05394](https://doi.org/10.48550/arXiv.1804.05394)
- [10] H. Abidi, R. Amami, and M. Pontier, Infinite horizon impulse control problem with continuous costs: numerical solutions, *Stochastics*, 89(6–7):1039–1060, 2017. doi: [10.1080/17442508.2017.1307985](https://doi.org/10.1080/17442508.2017.1307985).
- [11] H. Abidi, R. Amami, and M. Pontier, Infinite horizon impulse control problem with jumps and continuous switching costs. *Arab Journal of Mathematical Sciences*, 28(1):2–36, 2022. [10.1108/AJMS-10-2020-0088](https://doi.org/10.1108/AJMS-10-2020-0088).
- [12] J. Zhang, A numerical scheme for BSDEs, *Annals of Applied Probability*, 14(1):459–488, 2004. doi: [10.1214/aoap/1075828058](https://doi.org/10.1214/aoap/1075828058).
- [13] B. Bouchard and N. Touzi, Discrete-time approximation and Monte Carlo simulation of backward stochastic differential equations, *Stochastic Processes and their Applications*, 111(2):175–206, 2004. doi: [10.1016/j.spa.2004.01.001](https://doi.org/10.1016/j.spa.2004.01.001).
- [14] E. Gobet, J. P. Lemor, and X. Warin, A regression-based Monte Carlo method to solve backward stochastic differential equations, *Annals of Applied Probability*, 15(3):2172–2202, 2005. doi: [10.1214/105051605000000412](https://doi.org/10.1214/105051605000000412).
- [15] B. Bouchard and R. Elie, Discrete-time approximation of decoupled Forward–Backward SDE with jumps, *Stochastic Processes and their Applications*, 118(1):53–75, 2008. doi: [10.1016/j.spa.2007.03.010](https://doi.org/10.1016/j.spa.2007.03.010).
- [16] W. E, J. Han, and A. Jentzen, Deep learning-based numerical methods for high-dimensional parabolic partial Differential Equations and Backward Stochastic Differential Equations, *Communications in Mathematics and Statistics*, 5(4):349–380, 2017. doi: [10.1007/s40304-017-0117-6](https://doi.org/10.1007/s40304-017-0117-6).
- [17] J. Han, A. Jentzen, and W. E, Solving high-dimensional partial differential equations using deep learning, *PNAS*, 115(34):8505–8510, 2018. doi: [10.1073/pnas.1718942115](https://doi.org/10.1073/pnas.1718942115).
- [18] C. Beck, W. E, and A. Jentzen, Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations, *Journal of Nonlinear Science*, 29:1563–1619, 2019. doi: [10.1007/s00332-018-9525-3](https://doi.org/10.1007/s00332-018-9525-3).
- [19] C. Huré, H. Pham, and X. Warin, Deep backward schemes for high-dimensional nonlinear PDEs, *Mathematics of Computation*, 89(324):1547–1579, 2020. doi: [10.1090/mcom/3514](https://doi.org/10.1090/mcom/3514).
- [20] S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural Computation*, 9(8):1735–1780, 1997. doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- [21] A. Graves, A. Mohamed, and G. Hinton, Speech recognition with deep recurrent neural networks, In *ICASSP*, 2013. doi: [10.1109/ICASSP.2013.6638947](https://doi.org/10.1109/ICASSP.2013.6638947).
- [22] H. Abidi, BiLSTM-based approximation of backward stochastic differential equations, In *Proceedings of the IEEE International Conference on Recent Advances in Mathematics and Informatics (ICRAMI)*, 2025. doi: [10.1109/ICRAMI64946.2025.11472701](https://doi.org/10.1109/ICRAMI64946.2025.11472701).
- [23] I. Loshchilov and F. Hutter, Decoupled weight decay regularization, In *International Conference on Learning Representations (ICLR)*, 2017. URL: <https://api.semanticscholar.org/CorpusID:53592270>.

Author information

Hani Abidi, Department of Informatics and Applied Mathematics, Esprit School of Business LR11ES13, Research Laboratory for Mathematical Modeling, Statistical Reliability, and Stochastic Analysis, Tunisia.
E-mail: heni.abidi@esprit.tn