A Deep BSDE Method for a Class of Strongly Coupled FBSDEs

Christian Bender and Benedikt Simon Flierl

Department of Mathematics, Saarland University

September 14, 2026

We investigate a variant of the deep BSDE method introduced by E et al. [11, 13]. The key novelty is that we establish an a-posteriori convergence result for the approximation of strongly coupled forward-backward stochastic differential equations (FBSDEs), i.e., our result holds without any assumptions on small time horizons, monotonicity or weak coupling that are typically imposed in the literature on the deep BSDE method. Instead, we rely on smoothness assumptions on the coefficients and cover FBSDEs in which the coupling of the BSDE into the SDE depends on both the backward component Y and the control component Z. Numerical experiments illustrate the theoretical results and demonstrate the applicability of the proposed approach.

Keywords: deep BSDE, strongly coupled FBSDE, a-posteriori estimate, convergence
MSC 2020: 60H10, 60H30, 65C05, 65C30, 68T07

1. Introduction

In this paper, we study the numerical approximation of coupled forward-backward stochastic differential equations (FBSDEs) of the form

$$\begin{cases} X_t = x_0 + \displaystyle\int_0^t b(s, X_s, Y_s, Z_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s, Y_s)\,\mathrm{d}W_s, \\ Y_t = g(X_T) - \displaystyle\int_t^T F(s, X_s, Y_s, Z_s)\mathrm{d}s - \int_t^T Z_s\,\mathrm{d}W_s, \end{cases} \tag{1}$$

given deterministic functions $b\colon [0,T] \times \mathbb{R}^d \times \mathbb{R} \times (\mathbb{R}^d)^* \to \mathbb{R}^d$, $\sigma\colon [0,T] \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^{d\times d}$, $F\colon [0,T] \times \mathbb{R}^d \times \mathbb{R} \times (\mathbb{R}^d)^* \to \mathbb{R}$, $g\colon \mathbb{R}^d \to \mathbb{R}$ and an initial value $x_0 \in \mathbb{R}^d$, where $T \in \mathbb{R}_{>0}$ is a fixed time horizon and $d \in \mathbb{N}$. The equation is given on a complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, which carries a d-dimensional standard Brownian motion $(W_t)_{t\geq 0}$, such that $(\mathcal{F}_t)_{t\geq 0}$ is the natural filtration of $(W_t)_{t\geq 0}$ augmented with all the $\mathbb{P}$-null sets. Recall that a solution of the FBSDE (1) is a triple $(X, Y, Z) := (X_t, Y_t, Z_t)_{t\geq 0}$ of $(\mathcal{F}_t)_{t\geq 0}$-progressively measurable stochastic processes with values in $\mathbb{R}^d \times \mathbb{R} \times (\mathbb{R}^d)^*$, such that X and Y are continuous,

$\mathbb{E}\big[\sup_{t\in[0,T]}|X_t|^2\big], \mathbb{E}\big[\sup_{t\in[0,T]}|Y_t|^2\big], \int_0^T \mathbb{E}\big[|Z_t|^2\big] < \infty$, and, for any $t \in [0,T]$, FBSDE (1) holds $\mathbb{P}$-almost surely.

FBSDEs originate from the study of the Pontryagin maximum principle for stochastic control by Bismut [6] and Bensoussan [5], see also [31]. Besides several applications to mathematical finance [23, Chapter 8], FBSDEs are also closely connected to quasi-linear parabolic partial differential equations (PDEs) via non-linear Feynman-Kac representations [27, 28] and the Ma-Protter-Yong four-step scheme [20].

Well-posedness of FBSDEs can be established via fixed point iteration [2, 28] either for small time horizons or under certain restrictions on the type of the coupling (weak coupling) of X into (F, g) or on (Y, Z) into (b, σ), or under monotonicity assumptions on the coefficients. Alternatively, under sufficient smoothness and growth conditions on the coefficients, the FBSDE can be decoupled via the solution to the associated quasi-linear PDE, leading to well-posedness via the four-step scheme of Ma et al. [20]. The underlying idea of decoupling the equations finally led to the unifying approach to well-posedness of FBSDEs via decoupling fields by Ma et al. [22].

The close connection between FBSDEs and parabolic PDEs can also be exploited either way for the design of numerical schemes. E. g., by mimicking the four-step scheme, numerical techniques for PDEs can be made viable for the approximation of FBSDEs [10, 21], whereas directly approximating the FBSDEs and resorting to the Feynman-Kac representation lead to new probabilistic schemes for quasi-linear parabolic PDEs [9]. For a survey on numerical methods for BSDEs and FBSDEs, we refer to Chessari et al. [7].

In this paper, we focus on the deep BSDE method which has been introduced by E et al. [11, 13] for decoupled FBSDEs (i.e., for (b, σ) independent of (Y, Z)). We note that this method is closely related to the physics informed neural networks [29] in PDE numerics, see [26] for a comparison. The key idea of the deep BSDE method is to re-write the BSDE for Y in forward form as

$$Y_t = y_0 + \int_0^t F(s, X_s, Y_s, Z_s)\mathrm{d}s + \int_0^t Z_s\,\mathrm{d}W_s,$$

and to consider the control problem to choose (y_0, Z) in such a way that the terminal loss

$$\mathbb{E}[|Y_T - g(X_T)|^2]$$

becomes minimal. This idea can be traced back to Cvitanić and Zhang [8] who theoretically analyze the direction of steepest descent for this minimization problem in continuous time. The deep BSDE method [11, 13] makes this basic idea practical, by discretizing this control problem in time, parameterizing Z by a neural network, and applying stochastic gradient descent to minimize the terminal loss. In this approach, the terminal loss serves as a surrogate for the unobservable L^2-error between true solution and numerical approximation. Hence, in order to justify the deep BSDE method, a-posteriori estimates for controlling the true L^2-error in terms of the terminal loss are required. For decoupled FBSDEs, such a-posteriori estimates had already been available in the literature [3]. For coupled FBSDEs, the deep BSDE method has first been analyzed by Han and Long [14]. They consider the special case without Z-coupling and impose standard assumptions on a short time horizon or monotonicity or weak coupling (for brevity, we will refer to such a set of conditions as the weak coupling case from now on). Han and Long [14] derive suitable a-posteriori estimates for discrete-time versions of the coupled FBSDE system, which are finally combined with bounds for the time-discretization error due to Bender and Zhang [4]. Recently, the approach of Han and Long [14] has been extended to the case of

Z-coupling into the drift coefficient b by Negyesi et al. [25]; see Remark 3.2 for further discussion. Several variants and extensions of the deep BSDE method can be found in the literature (e.g., [12, 15, 30]), which typically rely on a weak coupling assumption in one way or another. A notable exception is the work by Jiang and Li [16], who assume sufficient smoothness conditions for the associated parabolic PDE to be well-posed, but consider the case in which coupling of the BSDE into the SDE is via Y and through the drift coefficient b only. Andersson et al. [1] provide a numerical test case with strong coupling, in which the deep BSDE method empirically fails to converge, and suggest a more robust variant of the deep BSDE method for FBSDEs arising in stochastic control. They attribute this failure of convergence to the Z-coupling: "This is only possible if the coupling of Z in b is strong enough, so that $g(X_T)$ can be efficiently controlled by Z" [1, Section 2.3, p. A232].

In this paper, we prove, to the best of our knowledge, the first a-posteriori estimates that justify the use of the deep BSDE method for FBSDEs with strong Z-coupling through the drift coefficient (and we, additionally, allow strong Y- and X-coupling in all coefficient functions). Similarly to Jiang and Li [16] we impose regularity and growth conditions on the coefficients which guarantee existence of a sufficiently 'nice' classical solution to the associated quasi-linear parabolic PDE. For the proof of our main result (Theorem 3.1), we identify an auxiliary decoupled FBSDE, to which we can apply the a-posteriori estimates of Bender and Steiner [3] and, then, analyze the error between the solutions to the auxiliary decoupled FBSDE and the original coupled FBSDE. Our results suggest that the failure of the deep BSDE method in the test case in [1] may rather be attributed to the combination of the quadratic growth of the coefficient function F with the strong Z-coupling. We also demonstrate our results numerically for a test case, in which the weak coupling conditions imposed in [25] are violated.

This paper is organized as follows. Section 2 introduces the framework and the standing assumptions. In Section 3, we discuss the proposed deep BSDE algorithm and state the main convergence theorem. Section 4 illustrates the method by numerical examples. The proof of the main result is given in Section 5. Section 6 presents easy-to-verify sufficient conditions on the coefficients ensuring the applicability of our results.

2. General Setting and Standing Assumptions

In what follows, we denote by $|\cdot|$ the Euclidean norm for (column and row) vectors and the Frobenius norm for matrices. When vectors are identified with single-column or single-row matrices, these norms coincides. Consequently, properties of the Frobenius norm, such as submultiplicativity, apply without further distinction to products of vectors and matrices.

We impose the following assumptions on b, σ, σ^{-1}, $\sigma\sigma^\top$, F, and g, where we formally define the functions σ^{-1} and $\sigma\sigma^\top$ as

$$\begin{aligned}\sigma^{-1}\colon [0,T]\times\mathbb{R}^d\times\mathbb{R}\to\mathbb{R}^{d\times d}, &\quad (t,x,y)\mapsto \sigma(t,x,y)^{-1},\\ \sigma\sigma^\top\colon [0,T]\times\mathbb{R}^d\times\mathbb{R}\to\mathbb{R}^{d\times d}, &\quad (t,x,y)\mapsto \sigma(t,x,y)\sigma(t,x,y)^\top.\end{aligned}$$

Assumption 2.1. *(i) The matrix inverse $\sigma^{-1}(t,x,y)$ of $\sigma(t,x,y)$ exists for all $(t,x,y)\in[0,T]\times\mathbb{R}^d\times\mathbb{R}$.*

(ii) *There exist constants* $H_b, H_\sigma, H_{\sigma^{-1}}, H_{\sigma\sigma^\top}, H_F, H_g \in \mathbb{R}_{>0}$ *such that for any two* (t,x,y,z), $(t',x',y',z') \in [0,T] \times \mathbb{R}^d \times \mathbb{R} \times (\mathbb{R}^d)^*$ *it holds that*

$$\begin{aligned}
|b(t,x,y,z) - b(t',x',y',z')| &\leq H_b\big(|t-t'|^{1/2} + |x-x'| + |y-y'| + |z-z'|\big),\\
|\sigma(t,x,y) - \sigma(t',x',y')| &\leq H_\sigma\big(|t-t'|^{1/2} + |x-x'| + |y-y'|\big),\\
|\sigma^{-1}(t,x,y) - \sigma^{-1}(t',x',y')| &\leq H_{\sigma^{-1}}\big(|t-t'|^{1/2} + |x-x'| + |y-y'|\big),\\
|(\sigma\sigma^\top)(t,x,y) - (\sigma\sigma^\top)(t',x',y')| &\leq H_{\sigma\sigma^\top}\big(|t-t'|^{1/2} + |x-x'| + |y-y'|\big),\\
|F(t,x,y,z) - F(t',x',y',z')| &\leq H_F\big(|t-t'|^{1/2} + |x-x'| + |y-y'| + |z-z'|\big),\\
|g(x) - g(x')| &\leq H_g|x-x'|,
\end{aligned}$$

i. e., the functions b *and* F *are* $\frac{1}{2}$*-Hölder continuous in* t *and Lipschitz continuous in* (x,y,z) *with constants* H_b *and* H_F*, the functions* σ, σ^{-1} *and* $\sigma\sigma^\top$ *are* $\frac{1}{2}$*-Hölder continuous in* t *and Lipschitz continuous in* (x,y) *with constants* H_σ, $H_{\sigma^{-1}}$ *and* $H_{\sigma\sigma^\top}$*, and the function* g *is Lipschitz continuous in* x *with constant* H_g.

(iii) *The functions* b, σ, σ^{-1}, *and* $\sigma\sigma^\top$ *are bounded by constants* $K_b, K_\sigma, K_{\sigma^{-1}}, K_{\sigma\sigma^\top} \in \mathbb{R}_{>0}$.

Remark 2.2. (i) Since the Frobenius norm is sub-multiplicative, the boundedness of σ already implies the boundedness of $\sigma\sigma^\top$.

(ii) In connection with the boundedness of σ, it is easily seen that the $\frac{1}{2}$-Hölder continuity in t and the Lipschitz continuity in (x,y) of σ are transferred to $\sigma\sigma^\top$. Indeed, for each two $(t,x,y), (t',x',y') \in [0,T] \times \mathbb{R}^d \times \mathbb{R}$, applying the triangle inequality and the sub-multiplicativity of the Frobenius norm, it follows that

$$\begin{aligned}
&|(\sigma\sigma^\top)(t,x,y) - (\sigma\sigma^\top)(t',x',y')|\\
&= |\sigma(t,x,y)\sigma(t,x,y)^\top - \sigma(t,x,y)\sigma(t',x',y')^\top\\
&\qquad + \sigma(t,x,y)\sigma(t',x',y')^\top - \sigma(t',x',y')\sigma(t',x',y')^\top|\\
&\leq |\sigma(t,x,y)||\sigma(t,x,y) - \sigma(t',x',y')| + |\sigma(t,x,y) - \sigma(t',x',y')||\sigma(t',x',y')|\\
&\leq 2K_\sigma H_\sigma\big(|t-t'|^{1/2} + |x-x'| + |y-y'|\big).
\end{aligned}$$

(iii) Similarly, but now relying on the boundedness of σ^{-1}, the $\frac{1}{2}$-Hölder continuity in t and the Lipschitz continuity in (x,y) of σ also carry over to σ^{-1}: For each two $(t,x,y), (t',x',y') \in [0,T] \times \mathbb{R}^d \times \mathbb{R}$, it holds that

$$\begin{aligned}
|\sigma^{-1}(t,x,y) - \sigma^{-1}(t',x',y')| &= \big|\sigma^{-1}(t',x',y')\big(\sigma(t',x',y') - \sigma(t,x,y)\big)\sigma^{-1}(t,x,y)\big|\\
&\leq |\sigma^{-1}(t',x',y')||\sigma(t',x',y') - \sigma(t,x,y)||\sigma^{-1}(t,x,y)|\\
&\leq K_{\sigma^{-1}}^2 H_\sigma\big(|t-t'|^{1/2} + |x-x'| + |y-y'|\big).
\end{aligned}$$

Remark 2.3. By standard results, it holds that the existence of σ^{-1} together with the boundedness of σ and σ^{-1} is equivalent to the uniform ellipticity condition

$$\forall (t,x,y) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}\colon \quad \nu I \leq \sigma(t,x,y)\sigma(t,x,y)^\top \leq \mu I, \tag{2}$$

where $\nu \leq \mu$ are some positive real constants and $I = \operatorname{diag}(1,\ldots,1) \in \mathbb{R}^{d\times d}$ denotes the identity matrix. Thus, together with Remark 2.2, it follows that Assumption 2.1 can be reduced to the Hölder and Lipschitz conditions on b, σ, F, and g, the boundedness condition on b, and the above uniform ellipticity condition.

With the following additional assumption we ensure that FBSDE (1) is uniquely solvable.

Assumption 2.4. *(a) The quasi-linear parabolic PDE*

$$\begin{cases} u_t(t,x) = F(t,x,u(t,x),v(t,x)) - \dfrac{1}{2}\mathrm{tr}\big[u_{xx}(t,x)(\sigma\sigma^\top)(t,x,u(t,x))\big] \\ \qquad\qquad - u_x(t,x)b(t,x,u(t,x),v(t,x)), \quad (t,x) \in [0,T]\times\mathbb{R}^d, \\ u(T,x) = g(x), \quad x \in \mathbb{R}^d, \end{cases} \tag{3}$$

where $v\colon [0,T]\times\mathbb{R}^d \to (\mathbb{R}^d)^*$ is defined via

$$v(t,x) \coloneqq u_x(t,x)\sigma(t,x,u(t,x)),$$

admits a classical solution $u\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ such that its derivatives u_x and u_{xx} are bounded by constants $K_{u_x} \in \mathbb{R}_{>0}$ and $K_{u_{xx}} \in \mathbb{R}_{>0}$. Here, considering $x = (x_1,\dots,x_d)^\top$, the derivatives u_t, u_x, and u_{xx} of u are given as

$$\begin{aligned} u_t(t,x) &\coloneqq \frac{\partial u(t,x)}{\partial t} \in \mathbb{R}, \\ u_x(t,x) &\coloneqq (u_{x_i})_{i=1,\dots,d} \coloneqq \left(\frac{\partial u(t,x)}{\partial x_i}\right)_{i=1,\dots,d} \in (\mathbb{R}^d)^*, \\ u_{xx}(t,x) &\coloneqq (u_{x_i x_j}(t,x))_{i,j=1,\dots,d} \coloneqq \left(\frac{\partial^2 u(t,x)}{\partial x_i \partial x_j}\right)_{i,j=1,\dots,d} \in \mathbb{R}^{d\times d}. \end{aligned}$$

(b) There exist constants $H_u, H_{u_x}, H_{u_{xx}} \in \mathbb{R}_{>0}$ and an $\alpha \in (0,1]$ such that the functions u, u_x and u_{xx} from (a) fulfill the following Hölder and Lipschitz conditions: For all $(t,x),(t',x') \in [0,T]\times\mathbb{R}^d$ it holds that

$$\begin{aligned} |u(t,x) - u(t',x')| &\le H_u\big(|t-t'|^{1/2} + |x-x'|\big), \\ |u_x(t,x) - u_x(t',x')| &\le H_{u_x}\big(|t-t'|^{1/2} + |x-x'|\big), \\ |u_{xx}(t,x) - u_{xx}(t',x')| &\le H_{u_{xx}}\big(|t-t'|^{\alpha/2} + |x-x'|^{\alpha}\big), \end{aligned}$$

i. e., u and u_x are $\frac{1}{2}$-Hölder continuous in t and Lipschitz continuous in x with constants H_u and H_{u_x} and u_{xx} is $\frac{\alpha}{2}$-Hölder continuous in t and α-Hölder continuous in x with constant $H_{u_{xx}}$.

Remark 2.5.

(a) Observe that $u_{xx}(t,x) = u_{xx}(t,x)^\top$ since for a classical solution u of PDE (3) the derivative u_{xx} is assumed to be continuous.

(b) The Lipschitz continuity of u and u_x in x already follows by the boundedness of u_x and u_{xx}, respectively.

Assumptions 2.1 and 2.4 will be in force throughout the whole paper without further notice.

Lemma 2.6. *There exist (minimal) constants $H_v, K_v \in \mathbb{R}_{\geq 0}$ with*

$$H_v \leq K_\sigma H_{u_x} + K_{u_x} H_\sigma (1 + H_u) \quad \text{and} \quad K_v \leq K_{u_x} K_\sigma$$

such that

(a) for all $(t, x), (t', x') \in [0, T] \times \mathbb{R}^d$ it holds that

$$|v(t, x) - v(t', x')| \leq H_v\big(|t - t'|^{1/2} + |x - x'|\big),$$

i. e., the function v is $\frac{1}{2}$-Hölder continuous in t and Lipschitz continuous in x with common constant H_v, and

(b) v is bounded by K_v.

Proof. Let $(t, x), (t', x') \in [0, T] \times \mathbb{R}^d$. The Hölder/Lipschitz condition follows from

$$\begin{aligned}
&|u_x(t, x)\sigma(t, x, u(t, x)) - u_x(t', x')\sigma(t', x', u(t', x'))| \\
&\leq |u_x(t, x)\sigma(t, x, u(t, x)) - u_x(t', x')\sigma(t, x, u(t, x))| \\
&\qquad + |u_x(t', x')\sigma(t, x, u(t, x)) - u_x(t', x')\sigma(t', x', u(t', x'))| \\
&\leq (K_\sigma H_{u_x} + K_{u_x} H_\sigma (1 + H_u))\big(|t - t'|^{1/2} + |x - x'|\big).
\end{aligned}$$

The boundedness condition follows directly by the sub-multiplicativity of the Frobenius norm from the boundedness of u_x and σ. □

Remark 2.7. Define

$$\vartheta \colon [0, T] \times \mathbb{R}^d \times \mathbb{R} \to (\mathbb{R}^d)^* \quad (t, x, y) \mapsto u_x(t, x)\sigma(t, x, y)$$

Similarly to Lemma 2.6, there exist constants $H_\vartheta, K_\vartheta \in \mathbb{R}_{>0}$ with

$$H_\vartheta \leq K_{u_x} H_\sigma + K_\sigma H_{u_x} \quad \text{and} \quad K_\vartheta \leq K_{u_x} K_\sigma$$

such that for all $(t, x, y), (t', x', y') \in [0, T] \times \mathbb{R}^d \times \mathbb{R}$ it holds that

$$|\vartheta(t, x, y) - \vartheta(t', x', y')| \leq H_\vartheta\big(|t - t'|^{1/2} + |x - x'| + |y - y'|\big)$$

and ϑ is bounded by K_ϑ.

Theorem 2.8. *FBSDE* (1) *admits a unique adapted solution (X, Y, Z). Especially, the processes Y and Z take the following form: For any $t \in [0, T]$ it holds that*

$$Y_t = u(t, X_t) \quad \text{and} \quad Z_t = v(t, X_t).$$

Proof. Follows from [23, Chapter 4, Theorem 1.1]. Revisiting the proof of [23, Chapter 4, Theorem 1.1], we observe that we do not need the uniform Lipschitz continuity in (x, y, z) of the mapping $(t, x, y, z) \mapsto z\sigma(t, x, y)$ but the uniform Lipschitz continuity of v in x (see Lemma 2.6) is already sufficient. □

Remark 2.9. To see that $Y_t = u(t, X_t)$ and $Z_t = v(t, X_t) = u_x(t, X_t)\sigma(t, X_t, u(t, X_t))$ indeed solve the backward stochastic differential equation for Y_t of FBSDE (1), it suffices to apply Itô's formula to $u(t, X_t)$ and insert the identities for $u_t(s, X_s)$ and $u(T, X_T)$ given by the quasi-linear parabolic PDE (3):

$$\begin{aligned}
Y_t &= u(T, X_T) - \int_t^T u_t(s, X_s) + u_x(s, X_s) b(s, X_s, Y_s, Z_s) \\
&\qquad + \frac{1}{2}\mathrm{tr}\big[u_{xx}(s, X_s)(\sigma\sigma^\top)(s, X_s, Y_s)\big]\,\mathrm{d}s - \int_t^T u_x(s, X_s)\sigma(s, X_s, Y_s)\,\mathrm{d}W_s. \\
&= g(X_T) - \int_t^T F(s, X_s, u(s, X_s), v(s, X_s)) - u_x(s, X_s) b(s, X_s, u(s, X_s), v(s, X_s)) \\
&\qquad - \frac{1}{2}\mathrm{tr}\big[u_{xx}(s, X_s)(\sigma\sigma^\top)(s, X_s, u(s, X_s))\big] + u_x(s, X_s) b(s, X_s, Y_s, Z_s) \\
&\qquad + \frac{1}{2}\mathrm{tr}\big[u_{xx}(s, X_s)(\sigma\sigma^\top)(s, X_s, Y_s)\big]\,\mathrm{d}s - \int_t^T u_x(s, X_s)\sigma(s, X_s, u(s, X_s))\,\mathrm{d}W_s \\
&= g(X_T) - \int_t^T F(s, X_s, Y_s, Z_s)\,\mathrm{d}s - \int_t^T Z_s\,\mathrm{d}W_s.
\end{aligned}$$

3. Approximation via a Deep BSDE Method

Let $\pi \coloneqq \{t_0, \ldots, t_N\} \subseteq [0, T]$ be a time grid with $0 =: t_0 < \ldots < t_N \coloneqq T$. Define the function $\Pi\colon [0, T] \to \pi$ by $\Pi(t) \coloneqq t_i$ for $t \in [t_i, t_{i+1})$, $i = 0, \ldots, N-1$, and $\Pi(T) \coloneqq t_N = T$. Assume for a moment that we already know the initial condition $Y_0 = u(0, x_0)$ of the process Y and the functional relationship $Z_t = v(t, X_t)$ while the functional relationship $Y_t = u(t, X_t)$ is still unknown. In this situation, one possible strategy to approximate the processes X and Y would be to reformulate FBSDE (1) as a system of forward SDEs only involving the processes X and Y and to apply the Euler-Maruyama method. Back in reality, where the necessary information is not given, we are only able to mimic the proposed strategy by instead searching for a tuple $\eta \coloneqq (y_0, \rho)$ of an initial value $y_0 \in \mathbb{R}$ and a candidate function $\rho\colon [0, T] \times \mathbb{R}^d \to (\mathbb{R}^d)^*$ such that the initial condition $Y_0 \approx y_0$ and the functional relationship $Z_t \approx \rho(t, X_t)$ hold approximately. Then, on $[0, T]$, we approximate the adapted solution (X, Y, Z) by the triple $(X^{\eta,\pi}, Y^{\eta,\pi}, Z^{\eta,\pi}) \coloneqq (X_t^{\eta,\pi}, Y_t^{\eta,\pi}, Z_t^{\eta,\pi})_{t\in[0,T]}$, which is determined via

$$\begin{cases}
X_{t_0}^{\eta,\pi} \coloneqq x_0,\ Y_{t_0}^{\eta,\pi} \coloneqq y_0, \\
Z_{t_i}^{\eta,\pi} \coloneqq \rho(t_i, X_{t_i}^{\eta,\pi}), \quad \forall i \in \{0, \ldots, N\}, \\
X_{t_{i+1}}^{\eta,\pi} \coloneqq X_{t_i}^{\eta,\pi} + b(t_i, X_{t_i}^{\eta,\pi}, Y_{t_i}^{\eta,\pi}, Z_{t_i}^{\eta,\pi})\Delta_i + \sigma(t_i, X_{t_i}^{\eta,\pi}, Y_{t_i}^{\eta,\pi})\Delta W_i, \quad \forall i \in \{0, \ldots, N-1\}, \\
Y_{t_{i+1}}^{\eta,\pi} \coloneqq Y_{t_i}^{\eta,\pi} + F(t_i, X_{t_i}^{\eta,\pi}, Y_{t_i}^{\eta,\pi}, Z_{t_i}^{\eta,\pi})\Delta_i + Z_{t_i}^{\eta,\pi}\Delta W_i, \quad \forall i \in \{0, \ldots, N-1\}, \\
X_t^{\eta,\pi} \coloneqq X_{\Pi(t)}^{\eta,\pi},\ Y_t^{\eta,\pi} \coloneqq Y_{\Pi(t)}^{\eta,\pi},\ Z_t^{\eta,\pi} \coloneqq Z_{\Pi(t)}^{\eta,\pi}, \quad \forall t \in [0, T] \setminus \pi,
\end{cases} \tag{4}$$

where, for $i \in \{0, \ldots, N-1\}$, $\Delta_i \coloneqq t_{i+1} - t_i \in \mathbb{R}_{>0}$ and $\Delta W_i \coloneqq W_{t_{i+1}} - W_{t_i} \in \mathbb{R}^d$.

To evaluate the quality of the approximating solution triple $(X^{\eta,\pi}, Y^{\eta,\pi}, Z^{\eta,\pi})$, i. e., to evaluate how close it is to the true adapted solution (X, Y, Z) on $[0, T]$, we are interested in the following L^2-approximation error:

$$\mathcal{E}_{\eta,\pi} \coloneqq \sup_{t\in[0,T]} \mathbb{E}\big[|X_t - X_t^{\eta,\pi}|^2\big] + \max_{t_i\in\pi} \mathbb{E}\big[|Y_{t_i} - Y_{t_i}^{\eta,\pi}|^2\big] + \mathbb{E}\bigg[\int_0^T |Y_t - Y_t^{\eta,\pi}|^2 + |Z_t - Z_t^{\eta,\pi}|^2\,\mathrm{d}t\bigg].$$

We consider the following function class:

$$\Theta_{K_\rho,H_\rho} \coloneqq \left\{ \rho\colon [0,T]\times\mathbb{R}^d \to (\mathbb{R}^d)^* \;\middle|\; \begin{array}{l} \rho \text{ is bounded by } K_\rho \in \mathbb{R}_{>0},\ \rho \text{ is } \tfrac{1}{2}\text{-Hölder continuous} \\ \text{in the first variable and Lipschitz continuous in the} \\ \text{second variable with common constant } H_\rho \in \mathbb{R}_{>0}, \\ \text{i. e., for all } (t,x),(t',x') \in [0,T]\times\mathbb{R}^d \text{ it holds that} \\ \quad |\rho(t,x)-\rho(t',x')| \le H_\rho\big(|t-t'|^{1/2}+|x-x'|\big). \end{array} \right\}.$$

It turns out that, if choosing $\rho \in \Theta_{K_\rho,H_\rho}$, the L^2-approximation error $\mathcal{E}_{\eta,\pi}$ can be bounded solely considering the expected squared difference $\mathbb{E}\big[|Y^{\eta,\pi}_{T_N} - g(X^{\eta,\pi}_{T_N})|^2\big]$, which measures how good the approximating processes $X^{\eta,\pi}_{T_N}$ and $Y^{\eta,\pi}_{T_N}$ fulfill the terminal condition, and the mesh size $|\pi| \coloneqq \max_{i=0,\dots,N-1}\Delta_i$.

Theorem 3.1. *Let $K_\rho, H_\rho \in \mathbb{R}_{>0}$. There exists some constant $C \in [1,\infty)$ such that, for every $\rho \in \Theta_{K_\rho,H_\rho}$, $y_0 \in \mathbb{R}$ and $|\pi| \le (C(1+K_\rho^4))^{-1}$, it holds that*

$$\mathcal{E}_{\eta,\pi} \le C\Big(\mathrm{e}^{CK_\rho^4}\mathbb{E}\big[|Y^{\eta,\pi}_T - g(X^{\eta,\pi}_T)|^2\big] + \mathrm{e}^{C(K_\rho^2+H_\rho^2)}(|\pi|+|\pi|^\alpha)\Big).$$

This theorem shows that, on $[0,T]$, the adapted solution (X,Y,Z) of FBSDE (1) can be approximated arbitrarily well by $(X^{\eta,\pi},Y^{\eta,\pi},Z^{\eta,\pi})$, provided that the candidate function ρ and the initial condition y_0 are chosen such that the terminal condition is satisfied in an L^2-optimal sense, i. e., $\mathbb{E}\big[|Y^{\eta,\pi}_T - g(X^{\eta,\pi}_T)|^2\big]$ is minimized over all $\eta \in \mathbb{R}\times\Theta_{K_\rho,H_\rho}$. Here, in view of Lemma 2.6, it is natural to choose $K_\rho \ge K_v$ and $H_\rho \ge H_v$, while avoiding unnecessarily large values of these constants. Otherwise, a good approximation may not be found, as the search is restricted to an inappropriate class of functions.

Remark 3.2. Han and Long [14] were the first to prove a-posteriori bounds of this type for coupled FBDEs. In their setting, the drift coefficient b does not depend on Z (i.e., no Z-coupling) and one of the following generic assumptions must be satisfied: (i) small time horizon (T small), (ii) Weak coupling of Y into the forward SDE (Lipschitz constants of b and σ in the y-variable small), (iii) weak coupling of X into the backward SDE (Lipschitz constants of F and g in the x-variable small), (iv) $(-F)$ strongly decreasing in the y-variable, (v) b strongly decreasing in the x-variable; see [14, Theorem 1'] for a precise statement. The results of [14] have been generalized by Negyesi et al. [25] to the case that Z couples into the forward SDE through the drift coefficient b. They additionally require a weak coupling condition of Z (Lipschitz constant of b in the z-variable small) in each of the cases (ii)–(v). We note that, by means of [28, Theorem 3.3], the weak coupling condition of Z is not required for the well-posedness of the FBSDE in cases (iv)–(v).

Remark 3.3. (a) Since we assumed g to be Lipschitz continuous with constant H_g, as long as FBSDE (1) is solvable, the following bound applies:

$$\begin{aligned}\mathbb{E}\big[|Y^{\eta,\pi}_T - g(X^{\eta,\pi}_T)|^2\big] &= \mathbb{E}\big[|Y^{\eta,\pi}_T - Y_T + g(X_T) - g(X^{\eta,\pi}_T)|^2\big] \\ &\le 2\Big(\mathbb{E}\big[|Y^{\eta,\pi}_T - Y_T|^2\big] + H_g^2\mathbb{E}\big[|X_T - X^{\eta,\pi}_T|^2\big]\Big) \\ &\le 2(1\vee H_g^2)\mathcal{E}_{\eta,\pi}.\end{aligned}$$

Hence, a small terminal loss $\mathcal{E}_{\eta,\pi}$ also is necessary for a good L^2-approximation of the true solution.

(b) In the case all constants from Assumptions 2.1 and 2.4 are explicitly known together with K_ρ and H_ρ, Equation (13) in the proof of Theorem 3.1 gives an explicit bound for $\mathcal{E}_{\eta,\pi}$ in terms of $\mathbb{E}\big[|Y_T^{\eta,\pi} - g(X_T^{\eta,\pi})|^2\big]$, $|\pi|$ and $|\pi|^\alpha$ whenever $|\pi|$ fulfills the condition of Theorem 5.9.

Based on Theorem 3.1, we propose a deep BSDE method to approximate the solution of FBSDE (1). The solution is approximated by the Euler-Maruyama scheme (4) on the equidistant time grid $\hat{\pi}_N := \{t_0, \dots, t_N\} \subseteq [0,T]$ defined by $t_i := \frac{i}{N}T$, $i = 0, \dots, N$, where the tuple $\eta = (y_0, \rho)$ of the initial value y_0 and the control function ρ is learned jointly by a neural network architecture $\hat{\eta} := (\hat{y}_0, \hat{\rho})$ such that the terminal mismatch is minimized. More precisely, the initial value y_0 is treated as trainable parameter $\hat{y}_0 \in \mathbb{R}$, while the function ρ is parameterized by a (residual) neural subnet $\hat{\rho}\colon \mathbb{R} \times \mathbb{R}^d \to (\mathbb{R}^d)^*$. Let

$$\hat{\Gamma} := \mathbb{R}^{w\times w} \times \mathbb{R}^{w\times w} \times \mathbb{R}^{w\times(1+d)} \times \mathbb{R}^{d\times w} \times \mathbb{R}^w \times \mathbb{R}^w \times \mathbb{R}^w \times \mathbb{R}^d$$

be the possible parameter space of $\hat{\rho}$ and

$$\hat{\gamma} := (W_1^r, W_2^r, W^i, W^o, b_1^r, b_2^r, b^i, b^o) \in \hat{\Gamma}$$

a concrete parameter combination. Given fixed values $w \in \mathbb{N}$ and $K_{\hat{\rho}} \in \mathbb{R}_{>0}$, the neural subnet $\hat{\rho}$ is defined as

$$\hat{\rho}\colon \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d, \quad (t,x) \mapsto h_{K_{\hat{\rho}}}\Bigg(W^o r\Bigg(\sigma\Bigg(W^i\begin{pmatrix} t \\ x \end{pmatrix} + b^i\Bigg)\Bigg) + b^o\Bigg), \tag{5}$$

with residual block

$$r\colon \mathbb{R}^w \to \mathbb{R}^w, \quad s \mapsto s + W_2^r \sigma(W_1^r \phi(s) + b_1^r) + b_2^r,$$

hyperbolic tangent activation function

$$\sigma\colon \mathbb{R}^w \to \mathbb{R}^w,\ (s_i)_{i=1,\dots,w} \mapsto (\tanh(s_i))_{i=1,\dots,w},$$

layer normalization

$$\phi\colon \mathbb{R}^w \to \mathbb{R}^w,\ (s_i)_{i=1,\dots,w} \mapsto \left(\frac{s_i - \frac{1}{w}\sum_{j=1}^w s_j}{\sqrt{\frac{1}{w}\sum_{j=1}^w (s_j - \frac{1}{w}\sum_{k=1}^w s_k)^2 + 0.001}}\right)_{i=1,\dots,w},$$

and final projection

$$h_{K_{\hat{\rho}}}\colon \mathbb{R}^d \to \mathbb{R}^d, \quad s \mapsto \begin{cases} s & \text{if } |s| \le K_{\hat{\rho}} \\ K_{\hat{\rho}} \frac{s}{|s|} & \text{if } |s| > K_{\hat{\rho}}. \end{cases}$$

onto the closed ball with radius $K_{\hat{\rho}}$. Thus, $\hat{\rho}$ is chosen from the space of functions

$$\hat{\Theta}_{K_{\hat{\rho}}} := \big\{\hat{\rho}\colon \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d \mid \hat{\rho} \text{ is of the form (5) with } \hat{\gamma} \in \hat{\Gamma}\big\}.$$

Since all network layers of the above defined neural subnets are Lipschitz continuous and the projection operator preserves Lipschitz continuity, any $\hat{\rho} \in \hat{\Theta}_{K_{\hat{\rho}}}$ is automatically Lipschitz and consequently, since we ensure the boundedness by the final projection $h_{K_{\hat{\rho}}}$, also satisfies the required Hölder and Lipschitz regularity condition. In particular:

$$\forall \hat{\rho} \in \hat{\Theta}_{K_{\hat{\rho}}} \colon \exists H_{\hat{\rho}} \in \mathbb{R}_{>0} \colon \hat{\rho} \in \Theta_{K_{\hat{\rho}}, H_{\hat{\rho}}},$$

i. e., any candidate control $\hat{\rho} \in \hat{\Theta}_{K_{\hat{\rho}}}$ fulfills the assumption of Theorem 3.1.

Theoretically, in view of Theorem 3.1, we want to find a joint network $\hat{\eta} := (\hat{y}_0, \hat{\rho})$ with parameters $(\hat{y}_0, \hat{\gamma})$ which solves the following minimization problem:

$$\min_{\hat{\eta} \in \mathbb{R} \times \hat{\Theta}_{K_{\hat{\rho}}}} \mathbb{E}\big[|Y_T^{\hat{\eta},\hat{\pi}_N} - g(X_T^{\hat{\eta},\hat{\pi}_N})|^2\big]. \tag{6}$$

In practice, we approximate this expectation by simulation: Given a set of $K \in \mathbb{N}$ realizations $\{\boldsymbol{\Delta} w^k := (\Delta w_0^k, \ldots, \Delta w_{N-1}^k)\}_{k=1,\ldots,K}$ of the Brownian increments $(\Delta W_0, \ldots, \Delta W_{N-1})$ under $\hat{\pi}_N$ and taking into account the corresponding K trajectories $\{(x^{\hat{\eta},\hat{\pi}_N,k}, y^{\hat{\eta},\hat{\pi}_N,k}, z^{\hat{\eta},\hat{\pi}_N,k})\}_{k=1,\ldots,K}$ of the process $(X^{\hat{\eta},\hat{\pi}_N}, Y^{\hat{\eta},\hat{\pi}_N}, Z^{\hat{\eta},\hat{\pi}_N})$, we define the loss function

$$\ell(\hat{y}_0, \hat{\gamma}; \{\boldsymbol{\Delta} w^k\}_{k=1,\ldots,K}) := \frac{1}{K} \sum_{k=1}^{K} \big|y_{t_N}^{\hat{\eta},\hat{\pi}_N,k} - g(x_{t_N}^{\hat{\eta},\hat{\pi}_N,k})\big|^2.$$

Then the training of the neural net $\hat{\eta}$ works as follows. Iteratively, starting from an initial neural net $\hat{\eta}^0 := (\hat{y}_0^0, \hat{\rho}^0)$ with parameters $(\hat{y}_0^0, \hat{\gamma}^0) \in \mathbb{R} \times \hat{\Gamma}$, the parameters are updated according to the standard Adam optimizer (with $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-7}$, $m_0 = v_0 = 0$) [17]. In particular, for each $l = 1, \ldots, L$ (where $L \in \mathbb{N}$), we first simulate K realizations $\{\boldsymbol{\Delta} w^{l,k} := (\Delta w_0^{l,k}, \ldots, \Delta w_{N-1}^{l,k})\}_{k=1,\ldots,K}$ of the Brownian increments $(\Delta W_0, \ldots, \Delta W_{N-1})$. Given the neural net $\hat{\eta}^{l-1} := (\hat{y}_0^{l-1}, \hat{\rho}^{l-1})$ with parameters $(\hat{y}_0^{l-1}, \hat{\gamma}^{l-1})$, we then compute the corresponding K trajectories $\{(x^{\hat{\eta}^{l-1},\hat{\pi}_N,k}, Y^{\hat{\eta}^{l-1},\hat{\pi}_N,k}, Z^{\hat{\eta}^{l-1},\hat{\pi}_N,k})\}_{k=1,\ldots,K}$ of the process $(X^{\hat{\eta}^{l-1},\hat{\pi}_N}, Y^{\hat{\eta}^{l-1},\hat{\pi}_N}, Z^{\hat{\eta}^{l-1},\hat{\pi}_N})$. Finally, we update the parameters according to

$$(\hat{y}_0^l, \hat{\gamma}^l) := (\hat{y}_0^{l-1}, \hat{\gamma}^{l-1}) - \lambda(l) \frac{\hat{m}_l}{\sqrt{\hat{v}_l} + \varepsilon},$$

where

$$\hat{m}_l := \frac{m_l}{1 - \beta_1^l} \quad \text{and} \quad \hat{v}_l := \frac{v_l}{1 - \beta_2^l},$$

with

$$\begin{aligned} m_l &:= \beta_1 m_{l-1} + (1 - \beta_1) G_l, \\ v_l &:= \beta_2 v_{l-1} + (1 - \beta_2) G_l^2, \\ G_l &:= \nabla_{(\hat{y}_0, \hat{\gamma})} \ell(\hat{y}_0^{l-1}, \hat{\gamma}^{l-1}; \{\boldsymbol{\Delta} w^{l,k}\}_{k=1,\ldots,K}), \end{aligned}$$

and, for initial learning rate $\lambda_0 \in \mathbb{R}_{>0}$, decay rate $k_\lambda \in \mathbb{R}_{>0}$ and decay steps n_λ,

$$\lambda \colon \{1, \ldots, L\} \to \mathbb{R}_{>0},\ l \mapsto \lambda_0 k_\lambda^{l/n_\lambda}$$

is a learning rate schedule. We denote the minimizing neural net of minimization problem (6) with

$$\hat{\eta}^* := (\hat{y}_0^*, \hat{\rho}^*) := (\hat{y}_0^L, \hat{\rho}^L),$$

which has parameters $(\hat{y}_0^*, \hat{\gamma}^*) := (\hat{y}_0^L, \hat{\gamma}^L)$ and depends on w, $K_{\hat{\rho}}$, N, K, L, $\hat{y}_0^0$, $\hat{\gamma}^0$, λ_0, k_λ, n_λ and $\{\boldsymbol{\Delta} w^{l,k}\}_{l=1,\ldots,L;\, k=1,\ldots,K}$.

Remark 3.4. (a) While the described algorithm controls the bound $K_{\hat{\rho}^*}$ of the optimizer $\hat{\rho}^*$, it does not explicitly enforce any bound on its Lipschitz constant $H_{\hat{\rho}^*}$. Consequently, a sequence $(H_{\hat{\rho}_N^*})$ of Lipschitz constants corresponding to a sequence of learned minimizers $(\hat{\rho}_N^*)$ may diverge as $N \to \infty$. Depending on the growth rate of the sequence $(H_{\hat{\rho}_N^*})$, this may affect convergence as the time discretization is refined. Nevertheless, such behavior was not observed in our numerical experiments.

(b) Simultaneous control of both the bound $K_{\hat{\rho}_N^*}$ and Lipschitz constant $H_{\hat{\rho}_N^*}$ of a minimizer $\hat{\rho}_N^*$ would in particular be beneficial in situations where the constants K_v and H_v are unknown. In this case, in order for the neural network to approximate plausible candidate controls, the admissible bounds $K_{\hat{\rho}_N^*}$ and $H_{\hat{\rho}_N^*}$ eventually need to exceed K_v and H_v. However, Theorem 3.1 shows that these constants cannot be allowed to grow arbitrarily fast as $N \to \infty$. Hence, to guarantee convergence of the deep BSDE method, their growth must be controlled appropriately. One possibility to additionally control the Lipschitz constant of $\hat{\rho}_N^*$ is, noticing that $\hat{\rho}$ is even differentiable, to penalize $\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}|J_{\hat{\rho}}(t,x)|_2$ exceeding H_ρ, i.e., for some $c \in \mathbb{R}_{\geq 0}$, solving the minimization problem

$$\min_{\hat{\eta}\in\mathbb{R}\times\hat{\Theta}_{K_{\hat{\rho}}}} \mathbb{E}\big[|Y_T^{\hat{\eta},\hat{\pi}_N} - g(X_T^{\hat{\eta},\hat{\pi}_N})|^2\big] + c\left(\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} |J_{\hat{\rho}}(t,x)|_2 - H_\rho\right)^+,$$

where $J_{\hat{\rho}}(t,x)$ denotes the Jacobian of $\hat{\rho}$ evaluated at (t,x), $|\cdot|_2$ denotes the spectral norm and $(\cdot)^+ := \max\{\cdot, 0\}$. In the learning process of the neural net $(\hat{y}_0, \hat{\rho})$, we could adjust the loss function accordingly as

$$\begin{aligned}
&\ell(\hat{y}_0, \hat{\gamma}; \{\boldsymbol{\Delta} w^k\}_{k=1,\dots,K}) \\
&:= \frac{1}{K}\sum_{k=1}^{K}\big|y_{t_N}^{\hat{\eta},\hat{\pi}_N,k} - g(x_{t_N}^{\hat{\eta},\hat{\pi}_N,k})\big|^2 + \frac{c}{\gamma}\log\left(\sum_{k=1}^{K}\sum_{i=0}^{N} \mathrm{e}^{\gamma\left(|J_{\hat{\rho}}(t_i,\hat{X}_{t_i}^{l,k})|_2 - H_\rho\right)^+}\right),
\end{aligned}$$

where $\gamma \in [1, \infty)$. This approximates the maximal violation of the Hölder constant.

(c) While in our approach ρ is approximated by a single neural network $\hat{\rho}$ which also takes t as an input, the original deep BSDE approach uses, for each $i = 0, \dots, N$, a separate neural network $\hat{\rho}_i$ to approximate $\rho(t_i, \cdot)$. However, the use of separate neural networks does not directly guarantee that the resulting approximation $\hat{\rho}$ satisfies the Hölder and Lipschitz conditions required in our framework. Moreover, deep BSDE methods in the literature typically rely on standard feed-forward neural networks rather than residual neural networks.

4. Numerical Experiments

Example 4.1. The following example is a modified version of [25, Example 1] such that b and σ fulfill the necessary boundedness conditions of Assumption 2.1. Let $\mathbf{1} := (1, \dots, 1)^\top \in \mathbb{R}^d$, I denote the identity matrix of dimension $d \times d$ and

$$u(t,x) = \mathrm{e}^{-r(T-t)}\sum_{i=1}^{d}\sin(x_i) + d + 1.$$

For $r, \bar{\sigma}, \kappa_y, \kappa_z \in \mathbb{R}_{>0}$, the coefficients of FBSDE (1) are given as

$$\begin{aligned} b(t,x,y,z) &= \kappa_y \bar{\sigma} T_y(y)\mathbf{1} + \kappa_z T_z(z)^\top, \\ \sigma(t,x,y) &= \bar{\sigma} T_y(y) I, \\ F(t,x,y,z) &= r(y-(d+1)) - \frac{1}{2}\mathrm{e}^{-r(T-t)}\bar{\sigma}^2 u(t,x)^2 \sum_{i=1}^d \sin(x_i) \\ &\qquad + \kappa_y \sum_{i=1}^d z_i + \kappa_z \mathrm{e}^{-2r(T-t)}\bar{\sigma} u(t,x) \sum_{i=1}^d \cos^2(x_i), \\ g(x) &= \sum_{i=1}^d \sin(x_i) + d + 1, \end{aligned}$$

where

$$T_y(y) := \begin{cases} 1, & y < 1, \\ y, & 1 \le y \le 2d+1, \\ 2d+1, & y > 2d+1, \end{cases} \quad \text{and} \quad T_z(z) := \begin{cases} z, & |z| \le \bar{\sigma}(2d+1)\sqrt{d}, \\ \bar{\sigma}(2d+1)\sqrt{d}\frac{z}{|z|}, & |z| > \bar{\sigma}(2d+1)\sqrt{d}. \end{cases}$$

Then, the solution of FBSDE (1) is

$$Y_t = u(t, X_t) \quad \text{and} \quad Z_t = v(t, X_t)$$

with

$$v(t,x) = \mathrm{e}^{-r(T-t)}\bar{\sigma} u(t,x) \begin{pmatrix} \cos(x_1) \\ \vdots \\ \cos(x_d) \end{pmatrix}.$$

It is easy to see that Assumptions 2.1 and 2.4 are satisfied with $\alpha = 1$.

In the following, we first consider the case $d = r = \bar{\sigma} = \kappa_y = \kappa_z = T = 1$ and $x_0 = \frac{\pi}{4}$. Then, the conditions of [25, Theorem 3] are not fulfilled (see also the following Remark 4.2). Nevertheless, the conditions of Theorem 3.1 hold and thus we should see convergence. We train neural networks $\eta_N^* = (\hat{y}_{0,N}^*, \hat{\rho}_N^*)$ for $N \in \{8, 16, 32, 64, 128, 256\}$. In each training, we sample $\hat{y}_0^0$ from $\mathrm{Unif}(0, 0.1)$ and each component of $\hat{\gamma}^0$ from $\mathcal{N}(0,1)$. We set the learning rate schedule by choosing $\lambda_0 = 0.01$, $k_\lambda = 0.9549926$ and $n_\lambda = 100$. The training process stops after $L = 5000$ iterations. We chose $K_{\hat{\rho}_N^*} = 6$. For each N, we repeat the training 50 times. The results are given in the following figures.

Figure 4.1 shows, for $t = 0.5$, the fitted ρ_N^*'s (blue) in comparison with the true v (black), grouped by the number of time steps N. Additionally, a histogram of 10^5 realizations of $X_{0.5}$ is shown in each subplot. As can be seen, in the region where $X_{0.5}$ has high probability mass, the fitted $\hat{\rho}_N^*$'s become increasingly close to the true v as the time grid is refined. The same holds for the fitted initial values $\hat{y}_{0,N}^*$: For each N, the squared differences to the true initial value $u(0, x_0)$, that is $|\hat{y}_{0,N}^* - u(0,x_0)|^2$, are displayed via boxplots in Figure 4.2. Lastly, Figure 4.3 compares the 5% quantile, median and 95% quantile of terminal errors $\mathbb{E}\big[|Y_T^{\hat{\eta}_N^*, \hat{\pi}_N} - g(X_T^{\hat{\eta}_N^*, \hat{\pi}_N})|^2\big]$ and the L^2-approximation error $\mathcal{E}_{\hat{\eta}_N^*, \hat{\pi}_N}$ in dependence of N. Additionally, the three components of $\mathcal{E}_{\hat{\eta}_N^*, \hat{\pi}_N}$ are given separately. All errors are calculated by averaging over the respective errors

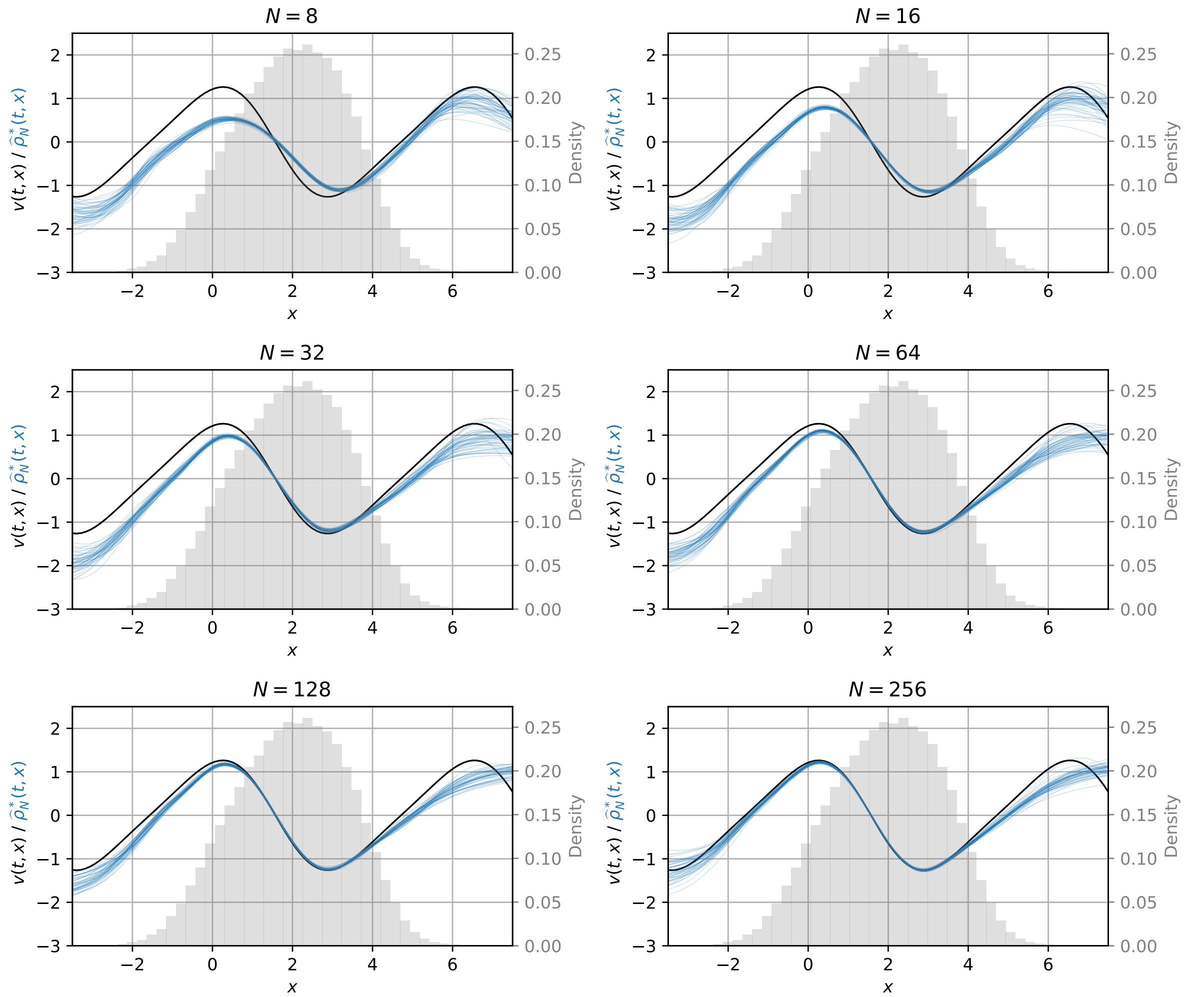

Figure 4.1: Fitted $\hat{\rho}_N^*$'s vs. true v for $t = 0.5$ grouped by the number of time steps N.

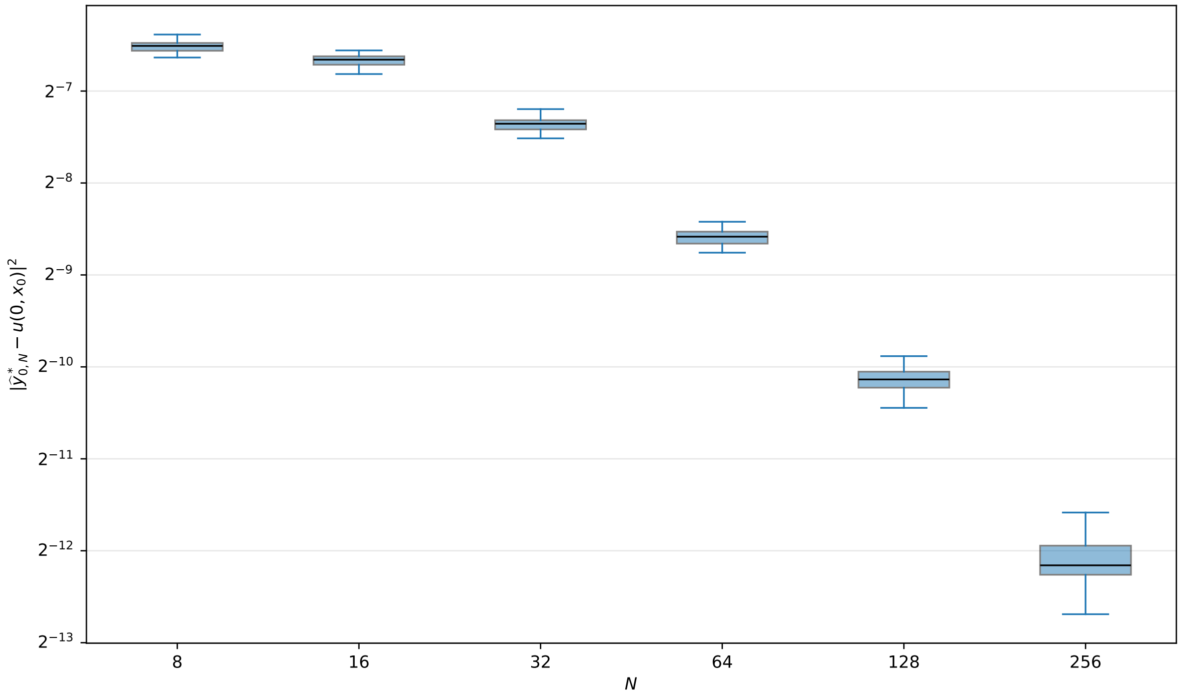

Figure 4.2: Boxplots of the errors $|\hat{y}^*_{0,N} - u(0, x_0)|^2$ grouped by the number of time steps N.

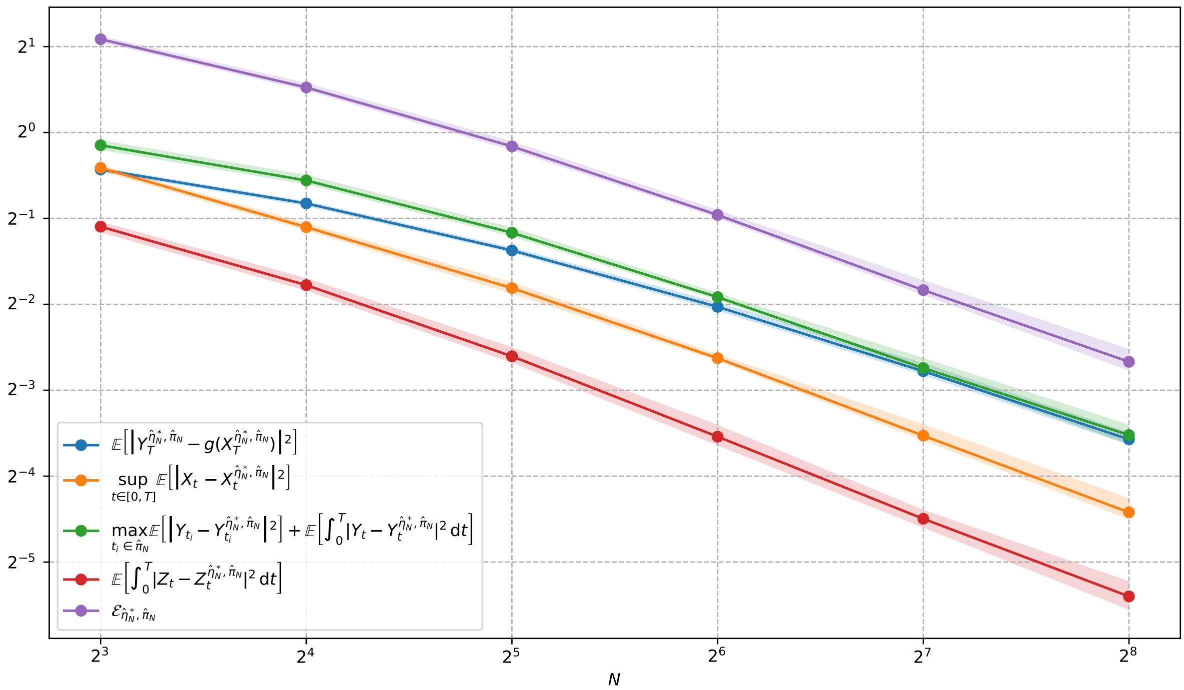

Figure 4.3: Medians of the respective errors (5%-95%-quantile bands shaded) for $d = 1$.

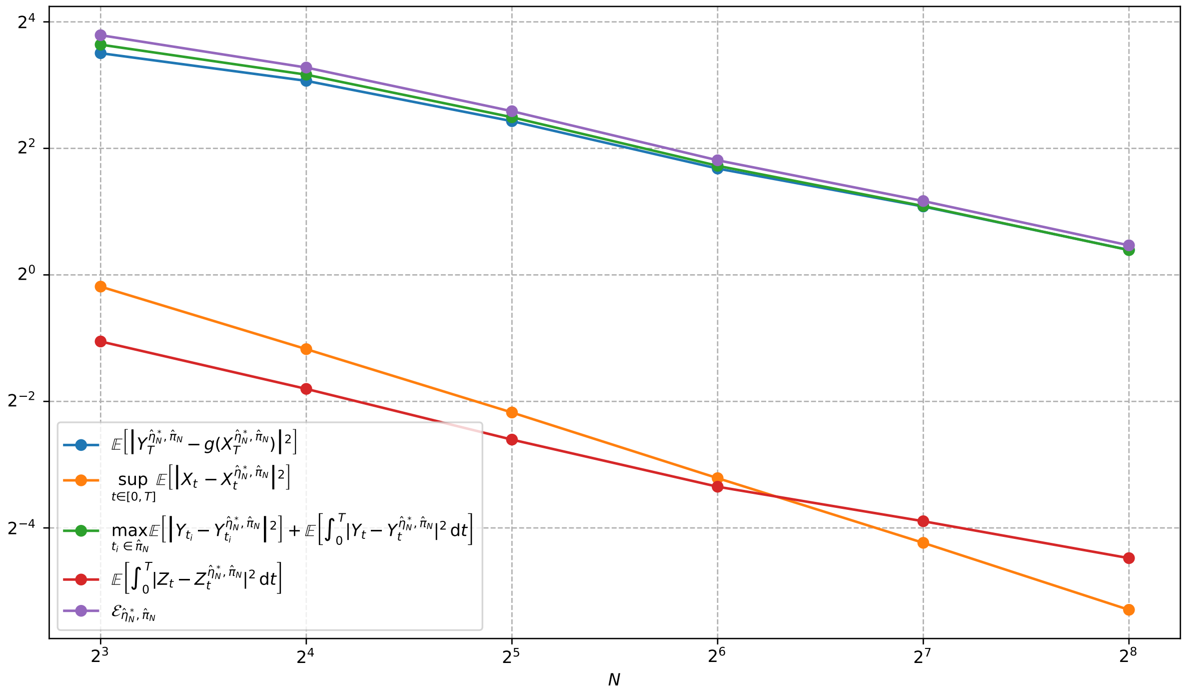

Figure 4.4: Errors for $d = 10$.

resulting from 25000 simulated independent paths of the Brownian motion. Overall, we nearly obtain an empirical convergence rate of $|\hat{\pi}_N|^{1/2}$, that is the convergence rate from Theorem 3.1.

Additionally, we consider the case $d = 10$, $r = 1$, $\bar{\sigma} = 0.2$, $\kappa_y = 0.2$, $\kappa_z = 0.1$, $T = 1$ and $x_0 = (\frac{\pi}{4}, \dots, \frac{\pi}{4})^\top$. Again, the conditions of [25, Theorem 3] are not fulfilled, but the conditions of Theorem 3.1 are. We train networks $\eta_N^* = (\hat{y}_{0,N}^*, \hat{\rho}_N^*)$ for $N \in \{8, 16, 32, 64, 128, 256\}$. In each training, we sample $\hat{y}_0^0$ from $\mathrm{Unif}(0, 0.1)$ and each component of $\hat{\gamma}^0$ from $\mathcal{N}(0, 1)$. We set the learning rate schedule by choosing $\lambda_0 = 0.01$, $k_\lambda = 0.99$ and $n_\lambda = 100$. The training process stops after $L = 2^{16}$ iterations. We choose $K_{\hat{\rho}_N^*} = 100$. The terminal errors $\mathbb{E}\big[|Y_T^{\hat{\eta}_N^*,\hat{\pi}_N} - g(X_T^{\hat{\eta}_N^*,\hat{\pi}_N})|^2\big]$ and the L^2-approximation error $\mathcal{E}_{\hat{\eta}_N^*,\hat{\pi}_N}$ including its three components are are given in Figure 4.4. Again, all errors are calculated by averaging over the respective errors resulting from 25000 simulated independent paths of the Brownian motion. The empirical convergence rate is slightly worse than $|\hat{\pi}_N|^{1/2}$.

Remark 4.2. For the given coefficients in the example above, it is a necessary condition for the applicability of [25, Theorem 3] that

$$\inf_{\lambda_1 \in \mathbb{R}_{>0}} \kappa_z^2 d^2 \frac{\mathrm{e}^{\lambda_1 \mathrm{T}}}{\lambda_1} = \kappa_z^2 d^2 T \mathrm{e} < 1.$$

This follows since, in the setting of [25, Theorem 3], a lower bound for $\bar{B}$ is easily derived as

$$\bar{B} > \kappa_z^2 d^2 \frac{\mathrm{e}^{\lambda_1 T}}{\lambda_1},$$

noting that the minimal squared Lipschitz constant of b in z equals κ_z^2, the minimal squared Lipschitz constant of g in x equals d, and b and σ are independent of x. Consequently, if

$$\kappa_z^2 d^2 T \mathrm{e} \geq 1,$$

then [25, Theorem 3] cannot be applied. In particular, this is the case for both parameter sets considered in our example.

5. Proof of Theorem 3.1

5.1. The Auxiliary Decoupled FBSDE

In this section we will prove Theorem 3.1. To this end, we introduce an auxiliary decoupled FBSDE where the backward part and the control of the solution triple are connected with the forward part (and the time) solely via the same functions u, u_x and σ as the components of the solution as the original coupled FBSDE. This auxiliary decoupled FBSDE is then approximated by the same Euler-Maruyama scheme as the original FBSDE.

In what follows, let $\tilde{x} \coloneqq (x^\top, \tilde{y})^\top$ and $\tilde{x}' \coloneqq (x'^\top, \tilde{y}')^\top$ with $x, x' \in \mathbb{R}^d$ and $\tilde{y}, \tilde{y}' \in \mathbb{R}$, such that $\tilde{x}, \tilde{x}' \in \mathbb{R}^{d+1}$. Fix $y_0 \in \mathbb{R}$, $K_\rho, H_\rho \in \mathbb{R}_{>0}$ and $\rho \in \Theta_{K_\rho, H_\rho}$. We introduce the coefficients of the forward SDE of the auxiliary FBSDE as

$$b^\rho \colon [0,T] \times \mathbb{R}^{d+1} \to \mathbb{R}^{d+1}, \qquad (t, \tilde{x}) \mapsto \begin{pmatrix} b(t, x, \tilde{y}, \rho(t,x)) \\ F(t, x, \tilde{y}, \rho(t,x)) \end{pmatrix},$$

and

$$\sigma^\rho \colon [0,T] \times \mathbb{R}^{d+1} \to \mathbb{R}^{(d+1)\times d}, \qquad (t, \tilde{x}) \mapsto \begin{pmatrix} \sigma(t, x, \tilde{y}) \\ \rho(t,x) \end{pmatrix},$$

Moreover, the generator $F^\rho \colon [0,T] \times \mathbb{R}^{d+1} \times \mathbb{R} \times (\mathbb{R}^d)^* \to \mathbb{R}$ of the backward SDE is given by

$$\begin{aligned} F^\rho(t, \tilde{x}, y, z) \coloneqq F(t, x, y, \zeta(t, \tilde{x}, y, z)) + u_x(t,x)\big[b(t, x, \tilde{y}, \rho(t,x)) - b(t, x, y, \zeta(t, \tilde{x}, y, z))\big] \\ + \frac{1}{2}\operatorname{tr}\big[u_{xx}(t,x)\big((\sigma\sigma^\top)(t, x, \tilde{y}) - (\sigma\sigma^\top)(t, x, y)\big)\big], \end{aligned}$$

where the function $\zeta \colon [0,T] \times \mathbb{R}^{d+1} \times \mathbb{R} \times (\mathbb{R}^d)^* \to \mathbb{R}$ is given as

$$\zeta(t, \tilde{x}, y, z) \coloneqq h(z)\sigma^{-1}(t, x, \tilde{y})\sigma(t, x, y)$$

with truncation

$$h \colon (\mathbb{R}^d)^* \to (\mathbb{R}^d)^*, \quad z \mapsto \begin{cases} z & \text{if } |z| \leq K_\rho \vee K_\vartheta, \\ (K_\rho \vee K_\vartheta)\frac{z}{|z|} & \text{if } |z| > K_\rho \vee K_\vartheta. \end{cases}$$

Obviously, the truncation h is bounded by $K_\rho \vee K_\vartheta$. Also, the truncation h is Lipschitz continuous with constant 1 as metric projection onto a closed convex set.

Instead of the original FBSDE (1), given the initial condition $\tilde{x}_0 := (x_0^\top, y_0)^\top$, we now look at the auxiliary FBSDE

$$\begin{cases} \tilde{X}_t^\eta = \tilde{x}_0 + \int_0^t b^\rho(s, \tilde{X}_s^\eta)\,\mathrm{d}s + \int_0^t \sigma^\rho(s, \tilde{X}_s^\eta)\,\mathrm{d}W_s, \\ Y_t^\eta = g(X_T^\eta) - \int_t^T F^\rho(s, \tilde{X}_s^\eta, Y_s^\eta, Z_s^\eta)\mathrm{d}s - \int_t^T Z_s^\eta\,\mathrm{d}W_s, \end{cases} \tag{7}$$

where Y_t^η and Z_t^η do not couple into the equation for $\tilde{X}_t^\eta := ((X_t^\eta)^\top, \tilde{Y}_t^\eta)^\top$.

We approximate the processes X^η, Y^η and Z^η by the triple $(X^{\eta,\pi}, Y^{\eta,\pi}, Z^{\eta,\pi})$, which we obtain from the Euler-Maruyama scheme (4). Additionally, to approximate the process $\tilde{Y}^\eta$, we define the process $\tilde{Y}^{\eta,\pi} \equiv Y^{\eta,\pi}$. Setting $\tilde{X}^{\eta,\pi} := ((X^{\eta,\pi})^\top, (\tilde{Y}^{\eta,\pi})^\top)^\top$, this 'extended' Euler-Maruyama scheme can easily be reformulated as

$$\begin{cases} \tilde{X}_{t_0}^{\eta,\pi} = \tilde{x}_0, \\ \tilde{X}_{t_{i+1}}^{\eta,\pi} = \tilde{X}_{t_i}^{\eta,\pi} + b^\rho(t_i, \tilde{X}_{t_i}^{\eta,\pi})\Delta_i + \sigma^\rho(t_i, \tilde{X}_{t_i}^{\eta,\pi})\Delta W_i, \ \forall i \in \{0, \ldots, N-1\}, \\ Z_{t_i}^{\eta,\pi} = \rho(t_i, X_{t_i}^{\eta,\pi}), \ \forall i \in \{0, \ldots, N\}, \\ \tilde{X}_t^{\eta,\pi} = \tilde{X}_{\Pi(t)}^{\eta,\pi}, \ Z_t^{\eta,\pi} = Z_{\Pi(t)}^{\eta,\pi}, \ \forall t \in [0,T] \setminus \pi, \\ Y_t^{\eta,\pi} = \tilde{Y}_t^{\eta,\pi}, \ \forall t \in [0,T]. \end{cases} \tag{8}$$

At this point it gets clear why the generator F^ρ is defined in such a way: We can reasonably use the 'same' approximation scheme to approximate the solutions of FBSDE (1) and FBSDE (7). In particular, at the grid points t_i, $i = 0, \ldots, N$, it holds that

$$\zeta(t_i, \tilde{X}_{t_i}^{\eta,\pi}, \tilde{Y}_{t_i}^{\eta,\pi}, \rho(t_i, X_{t_i}^{\eta,\pi})) = \rho(t_i, X_{t_i}^{\eta,\pi}).$$

Consequently, since $Y_{t_i}^{\eta,\pi} = \tilde{Y}_{t_i}^{\eta,\pi}$ and $Z_{t_i}^{\eta,\pi} = \rho(t_i, X_{t_i}^{\eta,\pi})$, we can simplify

$$\begin{aligned} F^\rho(t_i, \tilde{X}_{t_i}^{\eta,\pi}, Y_{t_i}^{\eta,\pi}, Z_{t_i}^{\eta,\pi}) &= F^\rho(t_i, \tilde{X}_{t_i}^{\eta,\pi}, \tilde{Y}_{t_i}^{\eta,\pi}, \rho(t_i, X_{t_i}^{\eta,\pi})) \\ &= F(t_i, X_{t_i}^{\eta,\pi}, \tilde{Y}_{t_i}^{\eta,\pi}, \rho(t_i, X_{t_i}^{\eta,\pi})) \\ &= F(t_i, X_{t_i}^{\eta,\pi}, Y_{t_i}^{\eta,\pi}, Z_{t_i}^{\eta,\pi}). \end{aligned} \tag{9}$$

Hence, for $i \in \{0, \ldots, N-1\}$, we can rewrite

$$Y_{t_{i+1}}^{\eta,\pi} = Y_{t_i}^{\eta,\pi} + F^\rho(t_i, \tilde{X}_{t_i}^{\eta,\pi}, Y_{t_i}^{\eta,\pi}, Z_{t_i}^{\eta,\pi})\Delta_i + Z_{t_i}^{\eta,\pi}\Delta W_i,$$

i. e., the tuple $(\tilde{X}^{\eta,\pi}, Y^{\eta,\pi}, Z^{\eta,\pi})$ can be considered as the deep BSDE approximation for the decoupled FBSDE with coefficients b^ρ, σ^ρ and F^ρ. We now first study this decoupled FBSDE and prove the following theorem.

Theorem 5.1. *FBSDE* (7) *admits a unique adapted solution* $(\tilde{X}^\eta, Y^\eta, Z^\eta)$*. Moreover, for each* $t \in [0,T]$*, it holds that*

$$Y_t^\eta = u(t, X_t^\eta) \quad \textit{and} \quad Z_t^\eta = \vartheta(t, X_t^\eta, \tilde{Y}_t^\eta).$$

To ensure that the FBSDE (7) is uniquely solvable certain regularity conditions concerning b^ρ, σ^ρ and F^ρ are crucial, which we will establish next. For later use, it is important to keep track of the dependence of the constants on K_ρ and H_ρ.

Remark 5.2. In what follows, we will repeatedly use the following basic inequalities without explicitly referring to them:

(i) Let $m \in \mathbb{N}$ and $a_1, \dots, a_m \in \mathbb{R}$. It holds that

$$\left(\sum_{i=1}^{m} a_i\right)^2 \leq m \sum_{i=1}^{m} a_i^2.$$

(ii) For any two $\tilde{x} = (x^\top, \tilde{y})^\top, \tilde{x}' = (x'^\top, \tilde{y}')^\top \in \mathbb{R}^{d+1}$, we have that

$$|x - x'| + |\tilde{y} - \tilde{y}'| \leq \sqrt{2}|\tilde{x} - \tilde{x}'|.$$

Lemma 5.3. *Let*

$$H_{b^\rho} := (\sqrt{2} + H_\rho)\sqrt{H_b^2 + H_F^2} \quad \textit{and} \quad H_{\sigma^\rho} := \sqrt{2H_\sigma^2 + H_\rho^2}.$$

For any $(t, \tilde{x}), (t', \tilde{x}') \in [0,T] \times \mathbb{R}^{d+1}$ it holds that

$$\begin{aligned} |b^\rho(t, \tilde{x}) - b^\rho(t', \tilde{x}')| &\leq H_{b^\rho}\big(|t - t'|^{1/2} + |\tilde{x} - \tilde{x}'|\big), \\ |\sigma^\rho(t, \tilde{x}) - \sigma^\rho(t', \tilde{x}')| &\leq H_{\sigma^\rho}\big(|t - t'|^{1/2} + |\tilde{x} - \tilde{x}'|\big), \end{aligned}$$

i. e., the functions b^ρ and σ^ρ are $\frac{1}{2}$-Hölder continuous in t and Lipschitz continuous in $\tilde{x}$ with constants H_{b^ρ} and H_{σ^ρ}.

Proof. The definition of the Euclidean norm in connection with the Hölder and Lipschitz conditions on b, F, and ρ imply

$$\begin{aligned} |b^\rho(t, \tilde{x}) - b^\rho(t', \tilde{x}')|^2 &= |b(t, x, \tilde{y}, \rho(t, x)) - b(t', x', \tilde{y}', \rho(t', x'))|^2 \\ &\quad + |F(t, x, \tilde{y}, \rho(t, x)) - F(t', x', \tilde{y}', \rho(t', x'))|^2 \\ &\leq (H_b^2 + H_F^2)\big(|t - t'|^{1/2} + |x - x'| + |\tilde{y} - \tilde{y}'| + |\rho(t, x) - \rho(t', x')|\big)^2 \\ &\leq (H_b^2 + H_F^2)(\sqrt{2} + H_\rho)^2\big(|t - t'|^{1/2} + |\tilde{x} - \tilde{x}'|\big)^2. \end{aligned}$$

Taking square roots on both sides shows the first inequality. Analogously, now taking the definition of the Frobenius norm into account and applying the $\frac{1}{2}$-Hölder and Lipschitz continutity of σ and ρ, we obtain

$$\begin{aligned} |\sigma^\rho(t, \tilde{x}) - \sigma^\rho(t, \tilde{x}')|^2 &= |\sigma(t, x, \tilde{y}) - \sigma(t', x', \tilde{y}')|^2 + |\rho(t, x) - \rho(t', x')|^2 \\ &\leq H_\sigma^2\big(|t - t'|^{1/2} + |x - x'| + |\tilde{y} - \tilde{y}'|\big)^2 + H_\rho^2\big(|t - t'|^{1/2} + |x - x'|\big)^2 \\ &\leq (2H_\sigma^2 + H_\rho^2)\big(|t - t'|^{1/2} + |\tilde{x} - \tilde{x}'|\big)^2. \end{aligned}$$

Taking square roots on both sides shows the second inequality. $\square$

Lemma 5.4. *Define*

$$L_{b^\rho} := \sqrt{12H_F^2 \vee \big(6H_F^2(T+K_\rho^2) + K_b^2 + 2F(0,0,0,0)^2\big)},$$
$$K_{\sigma^\rho} := \sqrt{K_\sigma^2 + K_\rho^2}.$$

For any $(t,\tilde{x}) \in [0,T] \times \mathbb{R}^{d+1}$ *it holds that*

$$|b^\rho(t,\tilde{x})|^2 \le L_{b^\rho}^2(1+|\tilde{x}|^2) \quad \text{and} \quad |\sigma^\rho(t,\tilde{x})|^2 \le K_{\sigma^\rho}^2.$$

Proof. Using the Hölder and Lipschitz conditions on F as well as the boundedness of b and ρ, we obtain

$$\begin{aligned}
|b^\rho(t,\tilde{x})|^2 &= |b(t,x,\tilde{y},\rho(t,x))|^2 + |F(t,x,\tilde{y},\rho(t,x))|^2 \\
&\le K_b^2 + 2|F(t,x,\tilde{y},\rho(t,x)) - F(0,0,0,0)|^2 + 2F(0,0,0,0)^2 \\
&\le K_b^2 + 2H_F^2\big(t^{1/2} + |x| + |\tilde{y}| + |\rho(t,x)|\big)^2 + 2F(0,0,0,0)^2 \\
&\le 12H_F^2|\tilde{x}|^2 + 6H_F^2(T+K_\rho^2) + K_b^2 + 2F(0,0,0,0)^2.
\end{aligned}$$

This shows the linear growth condition of b^ρ. The boundedness of σ^ρ follows directly from the boundedness of σ and ρ, i. e.,

$$|\sigma^\rho(t,\tilde{x})|^2 = |\sigma(t,x,\tilde{y})|^2 + |\rho(t,x)|^2 \le K_\sigma^2 + K_\rho^2. \qquad \square$$

Lemma 5.5. *Let* $\rho \in \Theta_{K_\rho,H_\rho}$. *Define*

$$\begin{aligned}
H_1 &:= (H_F + K_{u_x}H_b)\big(1 + H_\zeta^{(y,z)}\big) + \frac{1}{2}K_{u_{xx}}H_{\sigma\sigma^\top}, \\
H_2 &:= (H_F + K_{u_x}H_b)\big(1 + H_\zeta^{(t,x)}\big) + K_{u_x}H_b(\sqrt{2} + H_\rho) + 2K_bH_{u_x} + \frac{1+\sqrt{2}}{2}K_{u_{xx}}H_{\sigma\sigma^\top}, \\
H_3 &:= K_{\sigma\sigma^\top}H_{u_{xx}},
\end{aligned}$$

where

$$H_\zeta^{(y,z)} := K_{\sigma^{-1}}(K_\sigma \vee (K_\rho \vee K_\vartheta)H_\sigma) \quad \text{and} \quad H_\zeta^{(t,x)} := (K_\rho \vee K_\vartheta)(\sqrt{2}K_\sigma H_{\sigma^{-1}} + K_{\sigma^{-1}}H_\sigma).$$

(a) For all $(t,\tilde{x}) \in [0,T] \times \mathbb{R}^{d+1}$ *and any two* $(y,z),(y',z') \in \mathbb{R} \times (\mathbb{R}^d)^*$ *it holds that*

$$|F^\rho(t,\tilde{x},y,z) - F^\rho(t,\tilde{x},y',z')| \le H_1\big(|y-y'| + |z-z'|\big),$$

i. e., F^ρ *is Lipschitz in* (y,z) *uniformly in* $(t,\tilde{x})$ *with constant* H_1.

(b) For all $(y,z) \in \mathbb{R} \times (\mathbb{R}^d)^*$ *and any two* $(t,\tilde{x}),(t',\tilde{x}') \in [0,T] \times \mathbb{R}^{d+1}$ *it holds that*

$$|F^\rho(t,\tilde{x},y,z) - F^\rho(t',\tilde{x}',y,z)| \le H_2\big(|t-t'|^{1/2} + |\tilde{x}-\tilde{x}'|\big) + H_3\big(|t-t'|^{\alpha/2} + |\tilde{x}-\tilde{x}'|^\alpha\big).$$

Proof. (a) Using the shorthand notation

$$\tilde{\sigma}^{-1} := \sigma^{-1}(t,x,\tilde{y}), \qquad \sigma := \sigma(t,x,y), \qquad \sigma' := \sigma(t,x,y'),$$

we obtain

$$\begin{aligned}|\zeta(t,\tilde{x},y,z)-\zeta(t,\tilde{x},y',z')| &= |h(z)\tilde{\sigma}^{-1}\sigma - h(z')\tilde{\sigma}^{-1}\sigma'| \\ &= |(h(z)-h(z'))\tilde{\sigma}^{-1}\sigma + h(z')\tilde{\sigma}^{-1}(\sigma-\sigma')| \\ &\le K_\sigma K_{\sigma^{-1}}|z-z'| + (K_\rho \vee K_\vartheta)K_{\sigma^{-1}}H_\sigma|y-y'| \\ &\le H_\zeta^{(y,z)}\big(|y-y'|+|z-z'|\big).\end{aligned}$$

With this bound at hand, it easily follows that

$$\begin{aligned}|F(t,x,y,\zeta(t,\tilde{x},y,z)) - F(t,x,y',\zeta(t,\tilde{x},y',z'))| &\le H_F\big(1+H_\zeta^{(y,z)}\big)\big(|y-y'|+|z-z'|\big), \\ |b(t,x,y,\zeta(t,\tilde{x},y,z)) - b(t,x,y',\zeta(t,\tilde{x},y',z'))| &\le H_b\big(1+H_\zeta^{(y,z)}\big)\big(|y-y'|+|z-z'|\big).\end{aligned}$$

Thus, using the shorthand notation

$$b \coloneqq b(t,x,y,\zeta(t,\tilde{x},y,z)), \qquad b' \coloneqq b(t,x,y',\zeta(t,\tilde{x},y',z')), \qquad b_\rho \coloneqq b(t,x,\tilde{y},\rho(t,x)),$$

we have

$$\begin{aligned}|u_x(t,x)(b_\rho - b) - u_x(t,x)(b_\rho - b')| &= |u_x(t,x)||b'-b| \\ &\le K_{u_x}H_b\big(1+H_\zeta^{(y,z)}\big)\big(|y-y'|+|z-z'|\big).\end{aligned}$$

Lastly, it holds that

$$\begin{aligned}&\bigg|\frac{1}{2}\mathrm{tr}\big[u_{xx}(t,x)\big((\sigma\sigma^\top)(t,x,\tilde{y}) - (\sigma\sigma^\top)(t,x,y)\big)\big] \\ &\quad - \frac{1}{2}\mathrm{tr}\big[u_{xx}(t,x)\big((\sigma\sigma^\top)(t,x,\tilde{y}) - (\sigma\sigma^\top)(t,x,y')\big)\big]\bigg| \\ &= \frac{1}{2}\big|\mathrm{tr}\big[u_{xx}(t,x)\big((\sigma\sigma^\top)(t,x,y') - (\sigma\sigma^\top)(t,x,y)\big)\big]\big| \\ &\le \frac{1}{2}|u_{xx}(t,x)|\big|(\sigma\sigma^\top)(t,x,y') - (\sigma\sigma^\top)(t,x,y)\big| \\ &\le \frac{1}{2}K_{u_{xx}}H_{\sigma\sigma^\top}|y-y'|.\end{aligned}$$

The statement now follows by applying the triangle inequality to the difference $|F^\rho(t,\tilde{x},y,z) - F^\rho(t,\tilde{x},y',z')|$ and summing up the above bounds.

(b) Using the shorthand notation

$$\sigma \coloneqq \sigma(t,x,y), \qquad \sigma' \coloneqq \sigma(t',x',y), \qquad \tilde{\sigma}^{-1} \coloneqq \sigma^{-1}(t,x,\tilde{y}), \qquad (\tilde{\sigma}^{-1})' \coloneqq \sigma^{-1}(t',x',\tilde{y}'),$$

we obtain

$$\begin{aligned}|\zeta(t,\tilde{x},y,z)-\zeta(t',\tilde{x}',y,z)| &= |h(z)\tilde{\sigma}^{-1}\sigma - h(z)(\tilde{\sigma}^{-1})'\sigma'| \\ &= |h(z)|\big|\big(\tilde{\sigma}^{-1} - (\tilde{\sigma}^{-1})'\big)\sigma + (\tilde{\sigma}^{-1})'(\sigma-\sigma')\big| \\ &\le (K_\rho \vee K_\vartheta)(\sqrt{2}K_\sigma H_{\sigma^{-1}} + K_{\sigma^{-1}}H_\sigma)\big(|t-t'|^{1/2} + |\tilde{x}-\tilde{x}'|\big) \\ &= H_\zeta^{(t,x)}\big(|t-t'|^{1/2} + |\tilde{x}-\tilde{x}'|\big).\end{aligned}$$

Then, we have

$$|F(t,x,y,\zeta(t,\tilde{x},y,z)) - F(t',x',y,\zeta(t',\tilde{x}',y,z))| \le H_F\big(1+H_\zeta^{(t,x)}\big)\big(|t-t'|^{1/2}+|\tilde{x}-\tilde{x}'|\big),$$
$$|b(t,x,y,\zeta(t,\tilde{x},y,z)) - b(t',x',y,\zeta(t',\tilde{x}',y,z))| \le H_b\big(1+H_\zeta^{(t,x)}\big)\big(|t-t'|^{1/2}+|\tilde{x}-\tilde{x}'|\big).$$

Furthermore, it holds that

$$\begin{aligned}
&|b(t,x,\tilde{y},\rho(t,x)) - b(t',x',\tilde{y}',\rho(t',x'))| \\
&\le H_b\big(|t-t'|^{1/2}+|x-x'|+|\tilde{y}-\tilde{y}'|+|\rho(t,x)-\rho(t',x')|\big) \\
&\le H_b(\sqrt{2}+H_\rho)\big(|t-t'|^{1/2}+|\tilde{x}-\tilde{x}'|\big).
\end{aligned}$$

Now, using the shorthand notation

$$\begin{aligned}
b_\rho &:= b(t,x,\tilde{y},\rho(t,x)), & b &:= b(t,x,y,\zeta(t,\tilde{x},y,z)),\\
b'_\rho &:= b(t',x',\tilde{y}',\rho(t',x')), & b' &:= b(t',x',y,\zeta(t',\tilde{x}',y,z)),
\end{aligned}$$

we obtain

$$\begin{aligned}
&|u_x(t,x)(b_\rho-b) - u_x(t',x')(b'_\rho-b')| \\
&= |u_x(t,x)(b_\rho-b'_\rho+b'-b) + (u_x(t,x)-u_x(t',x'))(b'_\rho-b')| \\
&\le \big[K_{u_x}H_b\big(1+\sqrt{2}+H_\rho+H_\zeta^{(t,x)}\big)+2K_bH_{u_x}\big]\big(|t-t'|^{1/2}+|\tilde{x}-\tilde{x}'|\big).
\end{aligned}$$

Similarly, now abbreviating

$$\begin{aligned}
\sigma\sigma^\top &:= (\sigma\sigma^\top)(t,x,y), & \widetilde{\sigma\sigma}^\top &:= (\sigma\sigma^\top)(t,x,\tilde{y}),\\
(\sigma\sigma^\top)' &:= (\sigma\sigma^\top)(t',x',y), & (\widetilde{\sigma\sigma}^\top)' &:= (\sigma\sigma^\top)(t',x',\tilde{y}'),
\end{aligned}$$

we obtain

$$\begin{aligned}
&\left|\frac{1}{2}\mathrm{tr}\big[u_{xx}(t,x)\big(\widetilde{\sigma\sigma}^\top-\sigma\sigma^\top\big)\big] - \frac{1}{2}\mathrm{tr}\big[u_{xx}(t',x')\big((\widetilde{\sigma\sigma}^\top)'-(\sigma\sigma^\top)'\big)\big]\right| \\
&\le \frac{1}{2}\big|\mathrm{tr}\big[\big(u_{xx}(t,x)-u_{xx}(t',x')\big)\big(\widetilde{\sigma\sigma}^\top-\sigma\sigma^\top\big)\big]\big| \\
&\qquad + \frac{1}{2}\big|\mathrm{tr}\big[u_{xx}(t',x')\big(\widetilde{\sigma\sigma}^\top-(\widetilde{\sigma\sigma}^\top)'+(\sigma\sigma^\top)'-\sigma\sigma^\top\big)\big]\big| \\
&\le \frac{1}{2}|u_{xx}(t,x)-u_{xx}(t',x')||\widetilde{\sigma\sigma}^\top-\sigma\sigma^\top| + \frac{1}{2}|u_{xx}(t',x')||\widetilde{\sigma\sigma}^\top-(\widetilde{\sigma\sigma}^\top)'+(\sigma\sigma^\top)'-\sigma\sigma^\top| \\
&\le K_{\sigma\sigma^\top}H_{u_{xx}}\big(|t-t'|^{\alpha/2}+|\tilde{x}-\tilde{x}'|^\alpha\big) + \frac{1}{2}K_{u_{xx}}H_{\sigma\sigma^\top}(1+\sqrt{2})\big(|t-t'|^{1/2}+|\tilde{x}-\tilde{x}'|\big).
\end{aligned}$$

The statement now follows by applying the triangle inequality to the difference $|F^\rho(t,\tilde{x},y,z)-F^\rho(t',\tilde{x}',y,z)|$ and summing up the above bounds. □

Proof of Theorem 5.1. Since b^ρ and σ^ρ are uniformly Lipschitz in $\tilde{x}$ (Lemma 5.3) and fulfill linear growth conditions in $\tilde{x}$ (Lemma 5.4, for σ^ρ the boundedness implies the linear growth condition), the forward part of FBSDE (7) has a unique solution $\tilde{X}^\eta$ [24, Theorem 3.1 of Chapter 2]. In particular, $\tilde{X}^\eta$ is continuous and $(\mathcal{F}_t)_{t\ge0}$-adapted and consequently also $(\mathcal{F}_t)_{t\ge0}$-progressively measurable. Moreover, by [24, Lemma 3.2 of Chapter 2] in connection with Lemma 5.4, it holds that $\mathbb{E}\big[\sup_{t\in[0,T]}|\tilde{X}_t^\eta|^2\big] < \infty$. Thus, also the backward part admits a unique solution [23, Theorem 4.2 of Chapter 1] since F^ρ, considered as a function

$$F^\rho\colon [0,T]\times\Omega\times\mathbb{R}\times(\mathbb{R}^d)^* \to \mathbb{R}, \quad F^\rho(t,\omega,y,z) = F^\rho(t,\tilde{X}_t^\eta(\omega),y,z),$$

is Lipschitz in (y,z) uniformly in (t,ω) by Lemma 5.5(a) and satisfies

$$\begin{aligned}
&\mathbb{E}\bigg[|g(\tilde{X}_T^\eta)|^2 + \int_0^T |F^\rho(t,\tilde{X}_t^\eta,0,0)|^2\,\mathrm{d}t\bigg] \\
&\le 2H_g^2\mathbb{E}\Big[|\tilde{X}_T^\eta|^2\Big] + 2H_g^2 g(0)^2 + 2TF^\rho(0,0,0,0)^2 \\
&\qquad + 8\mathbb{E}\bigg[\int_0^T H_2^2 t + H_2^2|\tilde{X}_t^\eta|^2 + H_3^2 t^\alpha + H_3^2|\tilde{X}_t^\eta|^{2\alpha}\,\mathrm{d}t\bigg] \\
&< \infty
\end{aligned} \tag{10}$$

by the Lipschitz continuity of g, Lemma 5.5(b) and the integrability properties of $\tilde{X}^\eta$.

Now, fix some $t \in [0,T]$ and apply Itô's formula to $u(t,X_t^\eta)$. Additionally, using the identities for $u_t(x,X_s)$ and $u(T,X_T)$ given by the parabolic PDE (3) and noting that

$$\begin{aligned}
v(s,X_s^\eta) &= u_x(s,X_s^\eta)\sigma(s,X_s^\eta,u(s,X_s^\eta)) \\
&= h(\vartheta(s,X_s^\eta,\tilde{Y}_s^\eta))\sigma^{-1}(s,X_s^\eta,\tilde{Y}_s^\eta)\sigma(s,X_s^\eta,u(s,X_s^\eta)) \\
&= h(Z_s^\eta)\sigma^{-1}(s,X_s^\eta,\tilde{Y}_s^\eta)\sigma(s,X_s^\eta,Y_s^\eta) \\
&= \zeta(s,\tilde{X}_s^\eta,Y_s^\eta,Z_s^\eta),
\end{aligned}$$

since $\vartheta(s,X_s^\eta,\tilde{Y}_s^\eta)$ is bounded by $K_\vartheta \le K_\rho \vee K_\vartheta$, it follows that

$$\begin{aligned}
Y_t^\eta &= u(T,X_T^\eta) - \int_t^T u_s(s,X_s^\eta) + u_x(s,X_s^\eta)b(s,X_s^\eta,\tilde{Y}_s^\eta,\rho(s,X_s^\eta)) \\
&\qquad + \frac{1}{2}\mathrm{tr}\big[u_{xx}(s,X_s^\eta)(\sigma\sigma^\top)(s,X_s^\eta,\tilde{Y}_s^\eta)\big]\,\mathrm{d}s - \int_t^T u_x(s,X_s^\eta)\sigma(s,X_s^\eta,\tilde{Y}_s^\eta)\,\mathrm{d}W_s \\
&= g(X_T^\eta) - \int_t^T F(s,X_s^\eta,u(s,X_s^\eta),v(s,X_s^\eta)) \\
&\qquad\qquad + u_x(s,X_s^\eta)\big[b(s,X_s^\eta,\tilde{Y}_s^\eta,\rho(s,X_s^\eta)) - b(s,X_s^\eta,u(s,X_s^\eta),v(s,X_s^\eta))\big] \\
&\qquad + \frac{1}{2}\mathrm{tr}\big[u_{xx}(s,X_s^\eta)\big((\sigma\sigma^\top)(s,X_s^\eta,\tilde{Y}_s^\eta) - (\sigma\sigma^\top)(s,X_s^\eta,u(s,X_s^\eta))\big]\,\mathrm{d}s - \int_t^T Z_s^\eta\,\mathrm{d}W_s \\
&= g(X_T^\eta) - \int_t^T F(s,X_s^\eta,Y_s^\eta,\zeta(s,\tilde{X}_s^\eta,Y_s^\eta,Z_s^\eta)) \\
&\qquad\quad + u(s,X_s^\eta)\big[b(s,X_s^\eta,\tilde{Y}_s^\eta,\rho(s,X_s^\eta)) - b(s,X_s^\eta,Y_s^\eta,\zeta(s,\tilde{X}_s^\eta,Y_s^\eta,Z_s^\eta))\big] \\
&\qquad + \frac{1}{2}\mathrm{tr}\big[u_{xx}(s,X_s^\eta)\big((\sigma\sigma^\top)(s,X_s^\eta,\tilde{Y}_s^\eta) - (\sigma\sigma^\top)(s,X_s^\eta,Y_s^\eta)\big]\,\mathrm{d}s - \int_t^T Z_s^\eta\,\mathrm{d}W_s \\
&= g(X_T^\eta) - \int_t^T F^\rho(s,\tilde{X}_s^\eta,Y_s^\eta,Z_s^\eta)\,\mathrm{d}s - \int_t^T Z_s^\eta\,\mathrm{d}W_s.
\end{aligned}$$

That is Y^η and Z^η fulfill the backward SDE of FBSDE (7). Moreover, since u and ϑ fulfill Hölder and Lipschitz conditions and are consequently continuous, the processes Y^η and Z^η

are $(\mathcal{F}_t)_{t\geq 0}$-progressively measurable as functions of $\tilde{X}^\eta$. Also Y^η is continuous. Lastly, the necessary integrability conditions are fulfilled. Indeed, for Y^η, we obtain from the integrability properties of $\tilde{X}^\eta$ that

$$\mathbb{E}\left[\sup_{t\in[0,T]}|Y_t^\eta|^2\right] \leq 4H_u^2 T + 4H_u^2\mathbb{E}\left[\sup_{t\in[0,T]}|X_t^\eta|^2\right] + 2u(0,0)^2 < \infty.$$

For Z^η, the required integrability condition follows directly from the boundedness of ϑ. □

Lemma 5.6. *It holds that*

$$\mathbb{E}\left[\sup_{t\in[0,T]}|\tilde{X}_t^\eta|^2\right] \leq (1+3|\tilde{x}_0|^2)\mathrm{e}^{3(L_{b\rho}^2\vee K_{\sigma\rho}^2)T(T+4)} =: c_6.$$

Proof. Follows from [24, Lemma 3.2 of Chapter 2] in connection with Lemma 5.4. □

Lemma 5.7. *Let*

$$c_{11} \coloneqq 2(K_b^2|\pi| + K_\sigma^2).$$

It holds that

$$\sup_{t\in[0,T]}\mathbb{E}\big[|X_t^\eta - X_{\Pi(t)}^\eta|^2\big] \leq c_{11}|\pi|.$$

Proof. Let $t\in[0,T]$. Inserting the definitions of X_t^η and $X_{\Pi(t)}^\eta$ and applying Itô's isometry we obtain

$$\begin{aligned}\mathbb{E}\big[|X_t^\eta - X_{\Pi(t)}^\eta|^2\big] &\leq 2\mathbb{E}\left[\left|\int_{\Pi(t)}^t b(s,X_s^\eta,\tilde{Y}_s^\eta,\rho(s,X_s^\eta))\,\mathrm{d}s\right|^2\right] + 2\mathbb{E}\left[\left|\int_{\Pi(t)}^t \sigma(s,X_s^\eta,\tilde{Y}_s^\eta)\,\mathrm{d}W_s\right|^2\right]\\ &\leq 2(t-\Pi(t))^2K_b^2 + 2(t-\Pi(t))K_\sigma^2\\ &\leq 2(K_b^2|\pi| + K_\sigma^2)|\pi|.\end{aligned}$$
□

5.2. A-Posteriori Error of the Auxiliary FBSDE

At first, we derive an upper bound for the L^2-approximation error between $(\tilde{X}^\eta, Y^\eta, Z^\eta)$ and $(\tilde{X}^{\eta,\pi}, Y^{\eta,\pi}, Z^{\eta,\pi})$, that is

$$\mathcal{E}_\pi^\eta \coloneqq \sup_{t\in[0,T]}\mathbb{E}\big[|\tilde{X}_t^\eta - \tilde{X}_t^{\eta,\pi}|^2\big] + \max_{t_i\in\pi}\mathbb{E}\big[|Y_{t_i}^\eta - Y_{t_i}^{\eta,\pi}|^2\big] + \mathbb{E}\left[\int_0^T |Y_t^\eta - Y_t^{\eta,\pi}|^2 + |Z_t^\eta - Z_t^{\eta,\pi}|^2\,\mathrm{d}t\right],$$

which will rely on the a-posteriori estimates for BSDEs in [3]. Additionally, we will use the standard result that, under the usual Lipschitz and linear growth conditions, the Euler–Maruyama scheme converges strongly with order $\frac{1}{2}$; see, e.g., [18, Theorem 10.2.2]. To be able to indicate the dependence on K_ρ and H_ρ we give a self-contained proof of this result in Appendix A.

Lemma 5.8. *Define*

$$c_3 := 4\big(2T(TH_{b^\rho}^2 + H_{\sigma^\rho}^2) + L_{b^\rho}^2|\pi|(1+c_6) + K_{\sigma^\rho}^2\big)\mathrm{e}^{8T(TH_{b^\rho}^2+H_{\sigma^\rho}^2)}.$$

It holds that

$$\sup_{t\in[0,T]} \mathbb{E}\big[|\tilde{X}_t^\eta - \tilde{X}_t^{\eta,\pi}|^2\big] \le c_3|\pi|.$$

Proof. See Appendix A. □

Theorem 5.9. *Define*

$$\begin{aligned}
c_1 &:= 2(2T+1)(d_1 + dc_7),\\
c_4 &:= 2d_2\big(H_g^2 c_6 + g(0)^2\big) + 2(2T+1)\big(d_4 H_g^2 c_3 + 4d_5 H_2^2(1+c_3)T\big)\\
&\qquad + 4Td_3\big(H_F^2(T+2c_6) + F(0,0,0,0)^2 + 4K_{u_x}^2K_b^2 + K_{u_{xx}}^2K_{\sigma\sigma^\top}^2\big),\\
\tilde{c}_4 &:= 8(2T+1)d_5H_3^2(1+c_3^\alpha)T,\\
\Gamma &:= 4H_1^2(T+1)(d\vee 2) + 16TH_1^4(1+T)^2(d\vee 2)^2,
\end{aligned}$$

where

$$\begin{aligned}
c_7 &:= \left(\frac{6}{(d\vee 2)-1} + 2 + 16\big(2+4(1+T)TH_1^2(d\vee 2)\big)\right)\mathrm{e}^{\Gamma T},\\
d_1 &:= \left(\frac{d_4}{4}+2\right)(2+d)\big[2 + c_7\big(1+H_1^2T(T+d)\big)\big],\\
d_2 &:= 2H_1(2T+1)\tilde{d}_2 + 8(1+H_1^2T)\mathrm{e}^{(1+H_1)^2T} + 16H_1^2\mathrm{e}^{(1+H_1)^2T}(T+2+2H_1^2T)|\pi|,\\
d_3 &:= 2H_1(2T+1)\tilde{d}_3 + 8(1+H_1^2T)\mathrm{e}^{(1+H_1)^2T} + \big(16H_1^2\mathrm{e}^{(1+H_1)^2T}(3+2H_1^2T)+8\big)|\pi|,\\
d_4 &:= (1+2(1+2\beta_\pi)T)\mathrm{e}^{\alpha_\pi T} + 4 + 4\beta_\pi|\pi|,\\
d_5 &:= \frac{d_4}{4H_1},\\
\tilde{d}_2 &:= \frac{d_4}{4H_1}\Big[(4+4H_1^2)\mathrm{e}^{(1+H_1)^2T} + 8H_1^2\mathrm{e}^{(1+H_1)^2T}(T+2+2H_1^2T)|\pi|\Big],\\
\tilde{d}_3 &:= \frac{d_4}{4H_1}\Big[4(1+H_1^2T)\mathrm{e}^{(1+H_1)^2T} + \big(8H_1^2\mathrm{e}^{(1+H_1)^2T}(3+2H_1^2T)+4\big)|\pi|\Big],\\
\alpha_\pi &:= \beta_\pi\left(1 + \frac{1}{2(1-|\pi|/2)^2}|\pi|\right) + \frac{1}{2(1-|\pi|/2)^2},\\
\beta_\pi &:= (4H_1+16H_1^2)\mathrm{e}^{(4H_1+16H_1^2)|\pi|}.
\end{aligned}$$

For $|\pi| \le \min\{1, (8H_1(1+4H_1))^{-1}, \Gamma^{-1}\}$ *it holds that*

$$\begin{aligned}
&\max_{t_i\in\pi}\mathbb{E}\big[|Y_{t_i}^\eta - Y_{t_i}^{\eta,\pi}|^2\big] + \mathbb{E}\left[\int_0^T |Y_t^\eta - Y_t^{\eta,\pi}|^2 + |Z_t^\eta - Z_t^{\eta,\pi}|^2\,\mathrm{d}t\right]\\
&\le c_1\mathbb{E}\big[|g(X_T^{\eta,\pi}) - Y_T^{\eta,\pi}|^2\big] + c_4|\pi| + \tilde{c}_4|\pi|^\alpha.
\end{aligned}$$

Proof. We define the random variables

$$\xi := g(X_T^\eta) \quad \text{and} \quad \xi^\pi := g(X_{t_N}^{\eta,\pi}) = g(X_T^{\eta,\pi}),$$

as well as the random functions

$$f\colon \Omega \times [0,T] \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R} \quad \text{and} \quad f^\pi\colon \Omega \times \pi \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R},$$

with

$$\begin{aligned} f(t,y,z) &:= f(\omega,t,y,z) := F^\rho(t,\tilde{X}_t^\eta(\omega),y,z),\\ f^\pi(t_i,y,z) &:= f^\pi(\omega,t_i,y,z) := F^\rho(t_i,\tilde{X}_{t_i}^{\eta,\pi}(\omega),y,z). \end{aligned}$$

Since X_T^η and $X_T^{\eta,\pi}$ are $\mathcal{F}_T$-measurable, from the measurability of g it follows that ξ and ξ^π are $\mathcal{F}_T$-measurable. Furthermore, from the Lipschitz continuity of g we obtain

$$\begin{aligned} \mathbb{E}\big[|\xi|^2\big] &= \mathbb{E}\big[|g(X_T^\eta) - g(0) + g(0)|^2\big] \le 2H_g^2\mathbb{E}\big[|X_T^\eta|^2\big] + 2g(0)^2 < \infty,\\ \mathbb{E}\big[|\xi^\pi|^2\big] &= \mathbb{E}\big[|g(X_T^{\eta,\pi}) - g(0) + g(0)|^2\big] \le 2H_g^2\mathbb{E}\big[|X_T^{\eta,\pi}|^2\big] + 2g(0)^2 < \infty, \end{aligned}$$

i. e., ξ and ξ^π are square-intgrable. Since F^ρ is measurable (since it fulfills Hölder and Lipschitz conditions in all variables and is thus continuous) and $\tilde{X}_{t_i}^{\eta,\pi}$ is $\mathcal{F}_{t_i}$-measurable for all $t_i \in \pi$ it follows that f^π is measurable and, for every $t_i \in \pi$ and $(y,z) \in \mathbb{R} \times \mathbb{R}^d$, $f^\pi(t_i,y,z)$ is $\mathcal{F}_{t_i}$-measurable. Also, since F^ρ is measurable and $\tilde{X}^\eta$ is $(\mathcal{F}_t)_{t\ge 0}$-progressively measurable, f is measurable and, for every $(y,z) \in \mathbb{R}\times\mathbb{R}^d$, $f(\cdot,y,z)$ is $\mathcal{F}_t$-adapted. Moreover, from Lemma 5.5(a) it directly follows that f and f^π are Lipschitz in (y,z) uniformly in (ω,t) resp. (ω,t_i) with Lipschitz constant H_1. Lastly, as in the proof of Theorem 5.1, it holds that

$$\mathbb{E}\bigg[\int_0^T |f(t,0,0)|^2\,\mathrm{d}t\bigg] = \mathbb{E}\bigg[\int_0^T |F^\rho(t,\tilde{X}_t^\eta,0,0)\,\mathrm{d}t\bigg] \overset{(10)}{<} \infty,$$

and, similarly,

$$\begin{aligned} \mathbb{E}\big[|f^\pi(t_i,0,0)|^2\big] &= \mathbb{E}\big[|F^\rho(t_i,\tilde{X}_{t_i}^{\eta,\pi},0,0) - F^\rho(t_i,0,0,0) + F^\rho(t_i,0,0,0)|^2\big]\\ &\le 8\mathbb{E}\Big[H_2^2 t_i + H_2^2|\tilde{X}_{t_i}^{\eta,\pi}|^2 + H_3^2 t_i^\alpha + H_3^2|\tilde{X}_{t_i}^{\eta,\pi}|^{2\alpha}\Big] + 2F^\rho(0,0,0,0)^2\\ &< \infty. \end{aligned}$$

Thus, we can apply [3, Theorem 3.1], which yields

$$\begin{aligned} &\max_{t_i\in\pi}\mathbb{E}\big[|Y_{t_i}^\eta - Y_{t_i}^{\eta,\pi}|^2\big] + \mathbb{E}\bigg[\int_0^T |Y_t^\eta - Y_t^{\eta,\pi}|^2 + |Z_t^\eta - Z_t^{\eta,\pi}|^2\,\mathrm{d}t\bigg]\\ &\le 2(2T+1)(d_1+dc_7)\mathcal{E} + \bigg(d_2\mathbb{E}\big[|\xi|^2\big] + d_3\mathbb{E}\bigg[\int_0^T |F^\rho(r,\tilde{X}_r^\eta,0,0)|^2\,\mathrm{d}r\bigg]\bigg)|\pi|\\ &\quad + 2(2T+1)d_4\mathbb{E}\big[|\xi - \xi^\pi|^2\big] + 2(2T+1)d_5\int_0^T \mathbb{E}\big[|F^\rho(t,\tilde{X}_t^\eta,Y_t^\eta,Z_t^\eta) - F^\rho(\Pi(t),\tilde{X}_t^{\eta,\pi},Y_t^\eta,Z_t^\eta)|^2\big]\,\mathrm{d}t\\ &= 2(2T+1)(d_1+c_7)\mathcal{E} + (d_2(\mathrm{II}) + d_3(\mathrm{III}))|\pi| + 2(2T+1)d_4(\mathrm{IV}) + 2(2T+1)d_5(\mathrm{V}), \end{aligned}$$

where

$$\mathcal{E} := \mathbb{E}\big[|\xi^\pi - Y_{t_N}^{\eta,\pi}|^2\big] + \max_{1\le i\le N}\mathbb{E}\Bigg[\bigg|Y_{t_i}^{\eta,\pi} - Y_{t_0}^{\eta,\pi} - \sum_{j=0}^{i-1}\big(f^\pi(t_j,Y_{t_j}^{\eta,\pi},Z_{t_j}^{\eta,\pi})\Delta_j + Z_{t_j}^{\eta,\pi}\Delta W_j\big)\bigg|^2\Bigg].$$

Observe that

$$f^\pi(t_i, Y_{t_i}^{\eta,\pi}, Z_{t_i}^{\eta,\pi}) = F^\rho(t_i, \tilde{X}_{t_i}^{\eta,\pi}, Y_{t_i}^{\eta,\pi}, Z_{t_i}^{\eta,\pi}) \overset{(9)}{=} F(t_i, X_{t_i}^{\eta,\pi}, Y_{t_i}^{\eta,\pi}, \rho(t_i, X_{t_i}^{\eta,\pi})).$$

Consequently, since, by the recursion of the Euler-Maruyama scheme, it holds that

$$Y_{t_i}^{\eta,\pi} = Y_{t_0}^{\eta,\pi} + \sum_{j=0}^{i-1}\big(F(t_j, X_{t_j}^{\eta,\pi}, Y_{t_j}^{\eta,\pi}, \rho(t_j, X_{t_j}^{\eta,\pi}))\Delta_j + \rho(t_j, X_{t_j}^{\eta,\pi})\Delta W_j\big),$$

the term $\mathcal{E}$ simplifies to

$$\mathcal{E} = \mathbb{E}\big[|\xi^\pi - Y_{t_N}^{\eta,\pi}|^2\big].$$

Inserting the definition $\xi = g(X_T^\eta)$ and using the Lipschitz continuity of g, we get that

$$(\mathrm{II}) = \mathbb{E}\big[|g(X_T^\eta) - g(0) + g(0)|^2\big] \le 2\big(H_g^2\mathbb{E}\big[|X_T^\eta|^2\big] + g(0)^2\big) \le 2\big(H_g^2 c_6 + g(0)^2\big).$$

Similarly, now using the Hölder and Lipschitz conditions on F, it follows that

$$\begin{aligned}
(\mathrm{III}) \le 4\mathbb{E}\bigg[\int_0^T & |F(r, X_r^\eta, 0, 0) - F(0,0,0,0)|^2 + F(0,0,0,0)^2 \\
&+ \big|u_x(r, X_r^\eta)\big[b(r, X_r^\eta, \tilde{Y}_r^\eta, \rho(r, X_r^\eta)) - b(r, X_r^\eta, 0, 0)\big]\big|^2 \\
&+ \frac{1}{4}\mathrm{tr}\big[u_{xx}(r, X_r^\eta)\big((\sigma\sigma^\top)(r, X_r^\eta, \tilde{Y}_r^\eta) - (\sigma\sigma^\top)(r, X_r^\eta, 0)\big)\big]^2 \,\mathrm{d}r\bigg] \\
&\le 4H_F^2\left(2\int_0^T r + \mathbb{E}\big[|X_r^\eta|^2\big]\,\mathrm{d}r\right) + 4T\big(F(0,0,0,0)^2 + 4K_{u_x}^2K_b^2 + K_{u_{xx}}^2K_{\sigma\sigma^\top}^2\big) \\
&\le 4H_F^2\left(T^2 + 2T\mathbb{E}\bigg[\sup_{t\in[0,T]}|X_t^\eta|^2\bigg]\right) + 4T\big(F(0,0,0,0)^2 + 4K_{u_x}^2K_b^2 + K_{u_{xx}}^2K_{\sigma\sigma^\top}^2\big) \\
&\le 4T\big(H_F^2(T + 2c_6) + F(0,0,0,0)^2 + 4K_{u_x}^2K_b^2 + K_{u_{xx}}^2K_{\sigma\sigma^\top}^2\big).
\end{aligned}$$

Inserting the definitions $\xi = g(X_T^\eta)$ and $\xi^\pi = g(X_T^{\eta,\pi})$, using the Lipschitz continuity of g, and applying Lemma 5.8, we obtain

$$(\mathrm{IV}) \le H_g^2\mathbb{E}\big[|X_T^\eta - X_T^{\eta,\pi}|^2\big] \le H_g^2\mathbb{E}\big[|\tilde{X}_T^\eta - \tilde{X}_T^{\eta,\pi}|^2\big] \le H_g^2\sup_{t\in[0,T]}\mathbb{E}\big[|\tilde{X}_t^\eta - \tilde{X}_t^{\eta,\pi}|^2\big] \le H_g^2 c_3|\pi|.$$

From Lemma 5.5(b), Lemma 5.8 and Jensen's inequality we obtain

$$\begin{aligned}
(\mathrm{V}) &\le 2H_2^2\left(\int_0^T \mathbb{E}\Big[\big(|t - \Pi(t)|^{1/2} + |\tilde{X}_t^\eta - \tilde{X}_t^{\eta,\pi}|\big)^2\Big]\,\mathrm{d}t\right) \\
&\quad + 2H_3^2\left(\int_0^T \mathbb{E}\Big[\big(|t - \Pi(t)|^{\alpha/2} + |\tilde{X}_t^\eta - \tilde{X}_t^{\eta,\pi}|^\alpha\big)^2\Big]\,\mathrm{d}t\right) \\
&\le 4H_2^2\left(\int_0^T \mathbb{E}\big[|t - \Pi(t)|\big] + \sup_{s\in[0,T]}\mathbb{E}\big[|\tilde{X}_s^\eta - \tilde{X}_s^{\eta,\pi}|^2\big]\,\mathrm{d}t\right) \\
&\quad + 4H_3^2\left(\int_0^T \mathbb{E}\big[|t - \Pi(t)|^\alpha\big] + \sup_{s\in[0,T]}\mathbb{E}\big[|\tilde{X}_s^\eta - \tilde{X}_s^{\eta,\pi}|^{2\alpha}\big]\,\mathrm{d}t\right) \\
&\le 4H_2^2(1 + c_3)T|\pi| + 4H_3^2\left(\int_0^T \mathbb{E}\big[|t - \Pi(t)|^\alpha\big] + \left(\sup_{s\in[0,T]}\mathbb{E}\big[|\tilde{X}_s^\eta - \tilde{X}_s^{\eta,\pi}|^2\big]\right)^\alpha dt\right) \\
&\le 4H_2^2(1 + c_3)T|\pi| + 4H_3^2(1 + c_3^\alpha)T|\pi|^\alpha.
\end{aligned}$$

Putting the above estimates together gives the result. $\square$

5.3. Connecting the Original FBSDE and the Auxiliary Decoupled FBSDE

In a second step, we estimate the deviation between the adapted solution (X, Y, Z) of the original FBSDE (1) and the adapted solution (X^η, Y^η, Z^η) of the modified FBSDE (7):

$$\mathcal{E}_\eta := \sup_{t\in[0,T]} \mathbb{E}\big[|X_t - X_t^\eta|^2\big] + \max_{t_i\in\pi} \mathbb{E}\big[|Y_{t_i} - Y_{t_i}^\eta|^2\big] + \mathbb{E}\bigg[\int_0^T \big(|Y_t - Y_t^\eta|^2 + |Z_t - Z_t^\eta|^2\big)\,\mathrm{d}t\bigg].$$

Lemma 5.10. *Let*

$$\begin{aligned} c_8 &:= \Big[12H_b^2\big(1+(1+H_\vartheta)^2T\big)T + 10H_\sigma^2T\Big]c_1\mathrm{e}^{c_{10}},\\ c_9 &:= \Big[12H_b^2T^2\big((H_u+H_\rho+H_\vartheta H_u)^2 + H_u^2(1+H_\vartheta)^2c_{11} + (1+H_\rho+H_\vartheta)^2c_3\big)\\ &\qquad + 10H_\sigma^2T(H_u^2+c_3+H_u^2c_{11}) + \Big(12H_b^2\big(1+(1+H_\vartheta)^2T\big)T + 10H_\sigma^2T\Big)c_4\Big]\mathrm{e}^{c_{10}},\\ \tilde{c}_9 &:= \Big[12H_b^2\big(1+(1+H_\vartheta)^2T\big)T + 10H_\sigma^2T\Big]\tilde{c}_4\mathrm{e}^{c_{10}}, \end{aligned}$$

where

$$c_{10} := (1+H_u)^2T\big(12H_b^2(1+H_\vartheta)^2T + 10H_\sigma^2\big).$$

It holds that

$$\sup_{t\in[0,T]} \mathbb{E}\big[|X_t - X_t^\eta|^2\big] \le c_8\mathbb{E}\big[|\xi^\pi - Y_{t_N}^{\eta,\pi}|^2\big] + c_9|\pi| + \tilde{c}_9|\pi|^\alpha.$$

Proof. From the definitions of X_t and X_t^η, Hölder's inequality and Itô's isometry in connection with Fubini's theorem we obtain

$$\begin{aligned} \mathbb{E}\big[|X_t - X_t^\eta|^2\big] \le 2T\int_0^t \mathbb{E}\big[|b(s,X_s,Y_s,Z_s) - b(s,X_s^\eta,\tilde{Y}_s^\eta,\rho(s,X_s^\eta))|^2\big]\,\mathrm{d}s&\\ + 2\int_0^t \mathbb{E}\big[|\sigma(s,X_s,Y_s) - \sigma(s,X_s^\eta,\tilde{Y}_s^\eta)|^2\big]\,\mathrm{d}s.& \end{aligned}$$

Note that, using the identities of Theorem 5.1, we obtain

$$\begin{aligned} &|Y_s - \tilde{Y}_s^\eta|\\ &\le |Y_s - u(s,X_s^\eta)| + \big|u(s,X_s^\eta) - u\big(\Pi(s),X_{\Pi(s)}^\eta\big)\big| + \big|u\big(\Pi(s),X_{\Pi(s)}^\eta\big) - Y_{\Pi(s)}^{\eta,\pi}\big| + \big|Y_{\Pi(s)}^{\eta,\pi} - \tilde{Y}_s^\eta\big|\\ &= |u(s,X_s) - u(s,X_s^\eta)| + \big|u(s,X_s^\eta) - u\big(\Pi(s),X_{\Pi(s)}^\eta\big)\big| + \big|Y_{\Pi(s)}^\eta - Y_{\Pi(s)}^{\eta,\pi}\big| + |\tilde{Y}_s^{\eta,\pi} - \tilde{Y}_s^\eta|\\ &\le H_u|X_s - X_s^\eta| + H_u|s-\Pi(s)|^{1/2} + H_u\big|X_s^\eta - X_{\Pi(s)}^\eta\big| + \big|Y_{\Pi(s)}^\eta - Y_{\Pi(s)}^{\eta,\pi}\big| + |\tilde{X}_s^{\eta,\pi} - \tilde{X}_s^\eta|. \end{aligned} \tag{11}$$

Thus, together with Remark 2.7, we can bound

$$\begin{aligned} &|Z_s - \rho(s,X_s^\eta)|\\ &\le\ |Z_s - \vartheta(s,X_s^\eta,\tilde{Y}_s^\eta)| + \big|\vartheta(s,X_s^\eta,\tilde{Y}_s^\eta) - \rho\big(\Pi(s),X_{\Pi(s)}^{\eta,\pi}\big)\big| + \big|\rho\big(\Pi(s),X_{\Pi(s)}^{\eta,\pi}\big) - \rho(s,X_s^\eta)\big|\\ &=\ |\vartheta(s,X_s,Y_s) - \vartheta(s,X_s^\eta,\tilde{Y}_s^\eta)| + |Z_s^\eta - Z_s^{\eta,\pi}| + |\rho(\Pi(s),X_s^{\eta,\pi}) - \rho(s,X_s^\eta)|\\ &\le\ H_\vartheta|X_s - X_s^\eta| + H_\vartheta|Y_s - \tilde{Y}_s^\eta| + |Z_s^\eta - Z_s^{\eta,\pi}| + H_\rho|s-\Pi(s)|^{1/2} + H_\rho|X_s^\eta - X_s^{\eta,\pi}| \end{aligned}$$

$$
\begin{aligned}
&\overset{(11)}{\leq} H_\vartheta(1+H_u)|X_s - X_s^\eta| + |Z_s^\eta - Z_s^{\eta,\pi}| + (H_\rho + H_\vartheta H_u)|s-\Pi(s)|^{1/2} \\
&\qquad + (H_\rho + H_\vartheta)|\tilde{X}_s^\eta - \tilde{X}_s^{\eta,\pi}| + H_\vartheta H_u\big|X_s^\eta - X_{\Pi(s)}^\eta\big| + H_\vartheta\big|Y_{\Pi(s)}^\eta - Y_{\Pi(s)}^{\eta,\pi}\big|. \qquad (12)
\end{aligned}
$$

Then, exploiting the $\frac{1}{2}$-Hölder and Lipschitz continuity of b, σ, u, u_x and ρ as well as the boundedness of σ and u_x, it follows that

$$
\begin{aligned}
&\int_0^t \mathbb{E}\big[|b(s,X_s,Y_s,Z_s) - b(s,X_s^\eta,\tilde{Y}_s^\eta,\rho(s,X_s^\eta))|^2\big]\,\mathrm{d}s \\
&\leq H_b^2 \int_0^t \mathbb{E}\Big[\big(|X_s - X_s^\eta| + |Y_s - \tilde{Y}_s^\eta| + |Z_s - \rho(s,X_s^\eta)|\big)^2\Big]\,\mathrm{d}s \\
&\overset{(11),(12)}{\leq} H_b^2 \int_0^t \mathbb{E}\Big[\Big((1+H_u)(1+H_\vartheta)|X_s - X_s^\eta| + |Z_s^\eta - Z_s^{\eta,\pi}| + (H_u + H_\rho + H_\vartheta H_u)|s-\Pi(s)|^{1/2} \\
&\qquad + (1+H_\rho+H_\vartheta)|\tilde{X}_s^\eta - \tilde{X}_s^{\eta,\pi}| + H_u(1+H_\vartheta)\big|X_s^\eta - X_{\Pi(s)}^\eta\big| + (1+H_\vartheta)\big|Y_{\Pi(s)}^\eta - Y_{\Pi(s)}^{\eta,\pi}\big|\Big)^2\Big]\,\mathrm{d}s \\
&\leq 6H_b^2(1+H_u)^2(1+H_\vartheta)^2 \int_0^t \mathbb{E}\big[|X_s - X_s^\eta|^2\big]\,\mathrm{d}s \\
&\qquad + 6H_b^2\Big(\mathbb{E}\Big[\int_0^T |Z_s^\eta - Z_s^{\eta,\pi}|^2\Big]\mathrm{d}s + (1+H_\vartheta)^2 T \max_{t_i\in\pi} \mathbb{E}\big[|Y_{t_i}^\eta - Y_{t_i}^{\eta,\pi}|^2\big]\Big) \\
&\qquad + 6H_b^2 \int_0^T (H_u + H_\rho + H_\vartheta H_u)^2|\pi| + H_u^2(1+H_\vartheta)^2 \sup_{r\in[0,T]} \mathbb{E}\big[|X_r^\eta - X_{\Pi(r)}^\eta|^2\big] \\
&\qquad\qquad + (1+H_\rho+H_\vartheta)^2 \sup_{r\in[0,T]} \mathbb{E}\big[|\tilde{X}_r^\eta - \tilde{X}_r^{\eta,\pi}|^2\big]\,\mathrm{d}s \\
&\leq 6H_b^2(1+H_u)^2(1+H_\vartheta)^2 \int_0^t \mathbb{E}\big[|X_s - X_s^\eta|^2\big]\,\mathrm{d}s \\
&\qquad + 6H_b^2(1+(1+H_\vartheta)^2 T)\Big(c_1\mathbb{E}\big[|\xi^\pi - Y_{t_N}^{\eta,\pi}|^2\big] + c_4|\pi| + \tilde{c}_4|\pi|^\alpha\Big) \\
&\qquad + 6H_b^2 T\big((H_u + H_\rho + H_\vartheta H_u)^2 + H_u^2(1+H_\vartheta)^2 c_{11} + (1+H_\rho+H_\vartheta)^2 c_3\big)|\pi|.
\end{aligned}
$$

To derive the last inequality we additionally made use of Theorem 5.9 and Lemmas 5.7 and 5.8. Moreover, under the Lipschitz continuity of σ it holds that

$$
\begin{aligned}
&\int_0^t \mathbb{E}\big[|\sigma(s,X_s,Y_s) - \sigma(s,X_s^\eta,\tilde{Y}_s^\eta)|^2\big]\,\mathrm{d}s \\
&\leq H_\sigma^2 \int_0^t \mathbb{E}\Big[\big(|X_s - X_s^\eta| + |Y_s - \tilde{Y}_s^\eta|\big)^2\Big]\,\mathrm{d}s \\
&\overset{(11)}{\leq} 5H_\sigma^2 \int_0^t (1+H_u)^2\mathbb{E}\big[|X_s - X_s^\eta|^2\big] + H_u^2\mathbb{E}[|s-\Pi(s)|] + H_u^2\mathbb{E}\big[|X_s^\eta - X_{\Pi(s)}^\eta|^2\big] \\
&\qquad + \mathbb{E}\big[|Y_{\Pi(s)}^\eta - Y_{\Pi(s)}^{\eta,\pi}|^2\big] + \mathbb{E}\big[|\tilde{X}_s^{\eta,\pi} - \tilde{X}_s^\eta|^2\big]\,\mathrm{d}s \\
&\leq 5H_\sigma^2(1+H_u)^2 \int_0^t \mathbb{E}\big[|X_s - X_s^\eta|^2\big]\,\mathrm{d}s + 5H_\sigma^2 \int_0^t H_u^2|\pi| + H_u^2 \sup_{r\in[0,T]} \mathbb{E}\big[|X_r^\eta - X_{\Pi(r)}^\eta|^2\big] \\
&\qquad + \max_{t_i\in\pi} \mathbb{E}\big[|Y_{t_i}^\eta - Y_{t_i}^{\eta,\pi}|^2\big] + \sup_{r\in[0,T]} \mathbb{E}\big[|\tilde{X}_r^{\eta,\pi} - \tilde{X}_r^\eta|^2\big]\,\mathrm{d}s \\
&\leq 5H_\sigma^2(1+H_u)^2 \int_0^t \mathbb{E}\big[|X_s - X_s^\eta|^2\big] \\
&\qquad + 5H_\sigma^2 T\big(c_1\mathbb{E}\big[|\xi^\pi - Y_{t_N}^{\eta,\pi}|^2\big] + c_4|\pi| + \tilde{c}_4|\pi|^\alpha\big) + 5H_\sigma^2 T(H_u^2 + c_3 + H_u^2 c_{11})|\pi|.
\end{aligned}
$$

Together we obtain

$$\begin{aligned}\mathbb{E}\big[|X_t - X_t^\eta|^2\big] &\le \big[12H_b^2(1+H_u)^2(1+H_\vartheta)^2T + 10H_\sigma^2(1+H_u)^2\big]\int_0^t \mathbb{E}\big[|X_s - X_s^\eta|^2\big]\,\mathrm{d}s\\
&\quad + \big[12H_b^2\big(1+(1+H_\vartheta)^2T\big)T + 10H_\sigma^2T\big]\big(c_1\mathbb{E}\big[|\xi^\pi - Y_{t_N}^{\eta,\pi}|^2\big] + c_4|\pi| + \tilde{c}_4|\pi|^\alpha\big)\\
&\quad + \big[12H_b^2T^2\big((H_u+H_\rho+H_\vartheta H_u)^2 + H_u^2(1+H_\vartheta)^2c_{11} + (1+H_\rho+H_\vartheta)^2c_3\big)\\
&\qquad + 10H_\sigma^2T(H_u^2 + c_3 + H_u^2c_{11})\big]|\pi|.\end{aligned}$$

Applying Gronwall's lemma finishes the proof. □

Remark 5.11. From Equation (11) it follows that

$$\begin{aligned}\mathbb{E}\bigg[\int_0^t |Y_s - \tilde{Y}_s^\eta|^2\,\mathrm{d}s\bigg] &= \int_0^t \mathbb{E}\big[|Y_s - \tilde{Y}_s^\eta|^2\big]\,\mathrm{d}s\\
&\le 5\int_0^t H_u^2\mathbb{E}\big[|X_s - X_s^\eta|^2\big] + H_u^2\mathbb{E}\big[|s-\Pi(s)|\big] + H_u^2\mathbb{E}\big[|X_s^\eta - X_{\Pi(s)}^\eta|^2\big]\\
&\qquad + \mathbb{E}\big[|Y_{\Pi(s)}^\eta - Y_{\Pi(s)}^{\eta,\pi}|^2\big] + \mathbb{E}\big[|\tilde{X}_s^{\eta,\pi} - \tilde{X}_s^\eta|^2\big]\,\mathrm{d}s\\
&\le 5\int_0^t H_u^2\sup_{r\in[0,T]}\mathbb{E}\big[|X_r - X_r^\eta|^2\big] + H_u^2|\pi| + H_u^2\sup_{r\in[0,T]}\mathbb{E}\big[|X_r^\eta - X_{\Pi(r)}^\eta|^2\big]\\
&\qquad + \max_{t_i\in\pi}\mathbb{E}\big[|Y_{t_i}^\eta - Y_{t_i}^{\eta,\pi}|^2\big] + \sup_{r\in[0,T]}\mathbb{E}\big[|\tilde{X}_r^{\eta,\pi} - \tilde{X}_r^\eta|^2\big]\,\mathrm{d}s\\
&\le 5T(c_1 + H_u^2c_8)\mathbb{E}\big[|\xi^\pi - Y_{t_N}^{\eta,\pi}|^2\big] + 5T(\tilde{c}_4 + H_u^2\tilde{c}_9)|\pi|^\alpha\\
&\qquad + 5T\big(c_3 + c_4 + H_u^2(1+c_9+c_{11})\big)|\pi|.\end{aligned}$$

Theorem 5.12. *Define*

$$\begin{aligned}c_2 &:= \big(1 + H_u^2(1+T) + 2TH_\vartheta^2(1+5TH_u^2)\big)c_8 + 10T^2H_\vartheta c_1,\\
c_5 &:= \big(1 + H_u^2(1+T) + 2TH_\vartheta^2(1+5TH_u^2)\big)c_9 + 10T^2H_\vartheta\big(c_4 + c_3 + H_u^2(1+c_{11})\big),\\
\tilde{c}_5 &:= \big(1 + H_u^2(1+T) + 2TH_\vartheta^2(1+5TH_u^2)\big)\tilde{c}_9 + 10T^2H_\vartheta\tilde{c}_4.\end{aligned}$$

It holds that

$$\mathcal{E}_\eta \le c_2\mathbb{E}\big[|\xi^\pi - Y_{t_N}^{\eta,\pi}|^2\big] + c_5|\pi| + \tilde{c}_5|\pi|^\alpha.$$

Proof. Since $Y_t = u(t, X_t)$ and $Y_t^\eta = u(t, X_t^\eta)$, from the Lipschitz continuity of $u(t,x)$ in x uniformly in t and applying Lemma 5.10 it follows that

$$\begin{aligned}\max_{t_i\in\pi}\mathbb{E}\big[|Y_{t_i} - Y_{t_i}^\eta|^2\big] &= \max_{t_i\in\pi}\mathbb{E}\big[|u(t_i, X_{t_i}) - u(t_i, X_{t_i}^\eta)|^2\big]\\
&\le H_u^2\max_{t_i\in\pi}\mathbb{E}\big[|X_{t_i} - X_{t_i}^\eta|^2\big]\\
&\le H_u^2c_8\mathbb{E}\big[|\xi^\pi - Y_{t_N}^{\eta,\pi}|^2\big] + H_u^2c_9|\pi| + H_u^2\tilde{c}_9|\pi|^\alpha.\end{aligned}$$

Using the same arguments as above and applying Fubini's theorem, we also obtain

$$\mathbb{E}\bigg[\int_0^T |Y_t - Y_t^\eta|^2\,\mathrm{d}t\bigg] \le H_u^2Tc_8\mathbb{E}\big[|\xi^\pi - Y_{t_N}^{\eta,\pi}|^2\big] + H_u^2Tc_9|\pi| + H_u^2T\tilde{c}_9|\pi|^\alpha.$$

Similarly, now using the definitions $Z_t = v(t, X_t) = \vartheta(t, X_t, Y_t)$ and $Z_t^\eta = \vartheta(t, X_t^\eta, \tilde{Y}_t^\eta)$ as well as the Lipschitz continuity of $\vartheta(t, x, y)$ in (x, y) uniformly in t and the last remark, we get

$$\begin{aligned}
\mathbb{E}\Big[\int_0^T |Z_t - Z_t^\eta|^2 \,\mathrm{d}t\Big] &= \int_0^T \mathbb{E}\big[|\vartheta(t, X_t, Y_t) - \vartheta(t, X_t^\eta, \tilde{Y}_t^\eta)|^2\big] \,\mathrm{d}t \\
&\le 2H_\vartheta^2 \int_0^T \mathbb{E}\big[|X_t - X_t^\eta|^2\big] + \mathbb{E}\big[|Y_t - \tilde{Y}_t^\eta|^2\big] \,\mathrm{d}t \\
&\le 2TH_\vartheta^2\big(c_8 + 5T(c_1 + H_u^2 c_8)\big)\mathbb{E}\big[|\xi^\pi - Y_{t_N}^{\eta,\pi}|^2\big] \\
&\qquad + 2TH_\vartheta^2\big(\tilde{c}_9 + 5T(\tilde{c}_4 + H_u^2 \tilde{c}_9)\big)|\pi|^\alpha \\
&\quad + 2TH_\vartheta^2\big[c_9 + 5T\big(c_3 + c_4 + H_u^2(1 + c_9 + c_{11})\big)\big]|\pi|.
\end{aligned}$$

Summing up the three estimates and again applying Lemma 5.10 completes the proof. □

Proof of Theorem 3.1. It holds that

$$\mathcal{E}_{\eta,\pi} \le 2(\mathcal{E}_\eta + \mathcal{E}_\pi^\eta).$$

Thus, from Lemma 5.8, Theorem 5.9 and Theorem 5.12, it follows that

$$\mathcal{E}_{\eta,\pi} \le 2(c_1 + c_2)\mathbb{E}\big[|Y_T^{\eta,\pi} - g(X_T^{\eta,\pi})|^2\big] + 2(c_3 + c_4 + c_5)|\pi| + 2(\tilde{c}_4 + \tilde{c}_5)|\pi|^\alpha, \tag{13}$$

for $|\pi| \le \min\big\{1, (8H_1(1 + 4H_1))^{-1}, \Gamma^{-1}\big\}$. The statement now follows from some basic upper bounding of the constants in dependence of K_ρ and H_ρ (see Appendix B). For bounding β_π, note that $|\pi| \le (8H_1(1+4H_1))^{-1}$ by the assumption of Theorem 5.9, that is β_π can be bounded by a multiple of $4H_1 + 16H_1^2$. □

Remark 5.13. Carefully revisiting the proof of Theorem 3.1 one notices that the $|\pi|^\alpha$-part results from the u_{xx}-part of F^ρ. If σ is independent of Y_t, this u_{xx}-part vanishes and thus, in this case, the convergence in time holds up to order 1 instead of order α.

6. Sufficient Conditions on the Coefficient Functions

So far, the convergence does not fully rely on assumptions on the known functions b, σ, F, g and ρ, but also on assumptions on the unknown solution u of the parabolic PDE (3) together with its spatial derivatives u_x and u_{xx}. The goal of this section is to give additional conditions on b, σ, F and g, such that Assumption 2.4 is fulfilled. To this end, we make use of classical results on the well-posedness of parabolic Cauchy problems by Ladyženskaja et al. [19].

In the following we write

$$\begin{aligned}
x &= (x_1, \ldots, x_d)^\top, & b(t, x, y, z) &= (b_1(t, x, y, z), \ldots, b_d(t, x, y, z))^\top, \\
z &= (z_1, \ldots, z_d), & \sigma(t, x, y) &= (\sigma_{ij}(t, x, y))_{i,j=1,\ldots,d}.
\end{aligned}$$

Assumption 6.1. *(i) There exists an $\alpha \in (0, 1)$ such that $g \in H^{2+\alpha}(D)$ for any bounded domain $D \subset \mathbb{R}^d$ and g is bounded.*[1]

[1] For a definition of $H^{2+\alpha}(D)$, see [19, p. 7].

(ii) There exists a positive real constant $K_F \in \mathbb{R}_{>0}$ such that for all $(t,x) \in [0,T] \times \mathbb{R}^d$ it holds that

$$|F(t,x,0,0)| \leq K_F.$$

(iii) For all $i,j,k \in \{1,\dots,d\}$, the partial derivatives

$$\frac{\partial b_i}{\partial y}, \frac{\partial b_i}{\partial z_k}, \frac{\partial \sigma_{ij}}{\partial x_k}, \frac{\partial \sigma_{ij}}{\partial y}, \frac{\partial F}{\partial y}, \frac{\partial F}{\partial z_k}$$

exist, whereby

$$\frac{\partial \sigma_{ij}}{\partial x_k} \quad \text{and} \quad \frac{\partial \sigma_{ij}}{\partial y}$$

fulfill Hölder conditions in t, x and y with exponents $\frac{\alpha}{2}$, α and α.

Theorem 6.2. *Let Assumptions 2.1 and 6.1 be fulfilled. Then Assumption 2.4 holds true.*

Proof. For $(t,x,y,z) \in [0,T] \times \mathbb{R}^d \times \mathbb{R} \times (\mathbb{R}^d)^*$, define the functions (we set $\bar{t} := T - t$)

$$\begin{aligned}
a_i(t,x,y,z) &:= \frac{1}{2}\sum_{k=1}^{d} z_k \sum_{l=1}^{d} \sigma_{il}(\bar{t},x,y)\sigma_{lk}(\bar{t},x,y), \quad i = 1,\dots,d,\\
a(t,x,y,z) &:= \sum_{k=1}^{d} z_k b_k(\bar{t},x,y,z\sigma(\bar{t},x,y)) - F(\bar{t},x,y,z\sigma(\bar{t},x,y))\\
&\quad + \frac{1}{2}\sum_{i=1}^{d}\sum_{k=1}^{d} z_k \sum_{l=1}^{d}\left(\frac{\partial \sigma_{il}(\bar{t},x,y)}{\partial x_i}\sigma_{lk}(\bar{t},x,y) + \sigma_{il}(\bar{t},x,y)\frac{\partial \sigma_{lk}(\bar{t},x,y)}{\partial x_i}\right)\\
&\quad + \frac{1}{2}\sum_{i=1}^{d} z_i \sum_{k=1}^{d} z_k \sum_{l=1}^{d}\left(\frac{\partial \sigma_{il}(\bar{t},x,y)}{\partial y}\sigma_{lk}(\bar{t},x,y) + \sigma_{il}(\bar{t},x,y)\frac{\partial \sigma_{lk}(\bar{t},x,y)}{\partial y}\right),\\
a_{ij}(t,x,y,z) &:= \frac{1}{2}\sum_{l=1}^{d} \sigma_{il}(\bar{t},x,y)\sigma_{lj}(\bar{t},x,y), \quad i,j = 1,\dots,d,\\
A(t,x,y,z) &:= \sum_{l=1}^{d} z_l b_l(\bar{t},x,y,z\sigma(t,x,y)) - F(\bar{t},x,y,z\sigma(\bar{t},x,y)),\\
\psi_0(x) &:= g(x).
\end{aligned}$$

Note that a is well-defined, since, by Assumption 6.1(iii), the involved partial derivatives exist. By the same argument, for each $i,j \in \{1,\dots,d\}$, all of the following partial derivatives exist:

$$\begin{aligned}
\frac{\partial a_i(t,x,y,z)}{\partial x_j} &= \frac{1}{2}\sum_{k=1}^{d} z_k \sum_{l=1}^{d}\left(\frac{\partial \sigma_{il}(\bar{t},x,y)}{\partial x_j}\sigma_{lk}(\bar{t},x,y) + \sigma_{il}(\bar{t},x,y)\frac{\partial \sigma_{lk}(\bar{t},x,y)}{\partial x_j}\right),\\
\frac{\partial a_i(t,x,y,z)}{\partial y} &= \frac{1}{2}\sum_{k=1}^{d} z_k \sum_{l=1}^{d}\left(\frac{\partial \sigma_{il}(\bar{t},x,y)}{\partial y}\sigma_{lk}(\bar{t},x,y) + \sigma_{il}(\bar{t},x,y)\frac{\partial \sigma_{lk}(\bar{t},x,y)}{\partial y}\right),
\end{aligned}$$

$$\frac{\partial a_i(t,x,y,z)}{\partial z_j} = \frac{1}{2}\sum_{l=1}^{d} \sigma_{il}(\bar{t},x,y)\sigma_{lj}(\bar{t},x,y).$$

The above functions are defined in such a way that the identities

$$A(t,x,y,z) = a(t,x,y,z) - \sum_{i=1}^{d} \frac{\partial a_i(t,x,y,z)}{\partial x_i} - \sum_{i=1}^{d} \frac{\partial a_i(t,x,y,z)}{\partial y} z_i,$$
$$a_{ij}(t,x,y,z) = \frac{\partial a_i(t,x,y,z)}{\partial z_j}, \quad i,j = 1,\dots,d,$$

hold. Therefore, we are in the setting of [19, Chapter V, Theorem 8.1], which gives us sufficient conditions for the existence of a classical solution $\tilde{u} \in H^{1+\frac{\alpha}{2},2+\alpha}([0,T]\times\mathbb{R}^d)$[2] of the Cauchy problem

$$\begin{cases} \tilde{u}_t(t,x) = \sum_{i=1}^{d} \frac{\mathrm{d}}{\mathrm{d}x_i} a_i(t,x,\tilde{u}(t,x),\tilde{u}_x(t,x)) - a(t,x,\tilde{u}(t,x),\tilde{u}_x(t,x)), \quad (t,x)\in[0,T]\times\mathbb{R}^d, \\ \tilde{u}(0,x) = \psi_0(x), \quad x\in\mathbb{R}^d. \end{cases} \tag{14}$$

Note that a classical solution $\tilde{u} \in H^{1+\frac{\alpha}{2},2+\alpha}([0,T]\times\mathbb{R}^d)$ fulfills all boundedness, Hölder and Lipschitz conditions of Assumption 2.4.

In the following, we show that the above Cauchy problem (14) admits a classical solution $\tilde{u} \in H^{1+\frac{\alpha}{2},2+\alpha}([0,T]\times\mathbb{R}^d)$ by verifying the solvability conditions (a), (b), (c) of [19, Chapter V, Theorem 8.1]. In a final step, from $\tilde{u}$, we construct a solution u of the parabolic PDE (3) which inherits the desired boundedness, Hölder and Lipschitz conditions.

Solvability condition (a) coincides with Assumption 6.1(i). Solvability condition (b) follows from the uniform ellipticity condition (2) and the inequality

$$\begin{aligned} |A(t,x,y,0)y| &= |F(t,x,y,0)y| \\ &= |(F(t,x,y,0)-F(t,x,0,0))y + F(t,x,0,0)y| \\ &\le H_F|y|^2 + K_F|y| \\ &\le \left(H_F + \frac{1}{2}\right)|y|^2 + \frac{K_F^2}{2}, \end{aligned}$$

which uses the Lipschitz condition on F, Assumption 6.1(ii) and Young's inequality and holds true for arbitrary $(t,x,y,z)\in[0,T]\times\mathbb{R}^d\times\mathbb{R}\times(\mathbb{R}^d)^*$.

To verify solvability condition (c), first note that all functions

$$F,\ \sigma,\ b_k,\ \sigma_{ij},\ \frac{\partial\sigma_{ij}}{\partial x_k},\ \frac{\partial\sigma_{ij}}{\partial y}, \qquad i,j,k\in\{1,\dots,d\},$$

are $\frac{1}{2}$-Hölder continuous in t and Lipschitz continuous in x, y and z by Assumptions 2.1(ii) and 6.1(iii) and consequently continuous. Since additionally the mappings

$$(t,x,y,z)\mapsto \bar{t} = T-t \quad \text{and} \quad (t,x,y,z)\mapsto z_k, \quad k\in\{1,\dots,d\},$$

[2] For a definition of $H^{1+\frac{\alpha}{2},2+\alpha}([0,T]\times\mathbb{R}^d)$, see [19, pp. 7–8].

are obviously continuous, it follows that the functions

$$a_i,\ a,\ \frac{\partial a_i}{\partial x_j},\ \frac{\partial a_i}{\partial y},\ \frac{\partial a_i}{\partial z_j}, \qquad i,j\in\{1,\dots,d\},$$

are continuous.

Furthermore, for $|z|\le K_z\in\mathbb{R}_{>0}$ it holds that

$$\begin{aligned}
&|b_k(\bar t,x,y,z\sigma(\bar t,x,y)) - b_k(\bar t',x',y',z'\sigma(\bar t',x',y'))|\\
&\le |b(\bar t,x,y,z\sigma(\bar t,x,y)) - b(\bar t',x',y',z'\sigma(\bar t',x',y'))|\\
&\le H_b(1+K_zH_\sigma+K_\sigma)\big(|t-t'|^{1/2}+|x-x'|+|y-y'|+|z-z'|\big)
\end{aligned}$$

and

$$\begin{aligned}
&|F(\bar t,x,y,z\sigma(\bar t,x,y)) - F(\bar t',x',y',z'\sigma(\bar t',x',y'))|\\
&\le H_b(1+K_zH_\sigma+K_\sigma)\big(|t-t'|^{1/2}+|x-x'|+|y-y'|+|z-z'|\big),
\end{aligned}$$

i. e., the mappings $(t,x,y,z)\mapsto b_k(\bar t,x,y,z\sigma(\bar t,x,y))$ and $(t,x,y,z)\mapsto F(\bar t,x,y,z\sigma(\bar t,x,y))$ are $\frac{1}{2}$-Hölder continuous in t and Lipschitz continuous in x, y and z. The same also holds, obviously, for the projections $(t,x,y,z)\mapsto z_k$, $k\in\{1,\dots,d\}$, and, by Assumptions 2.1(ii) and 6.1(iii), for the mappings

$$(t,x,y,z)\mapsto\sigma_{ij}(\bar t,x,y),\ \frac{\partial\sigma_{ij}}{\partial x_k}(\bar t,x,y),\ \frac{\partial\sigma_{ij}}{\partial y}(\bar t,x,y),\quad i,j,k\in\{1,\dots,d\},$$

noting that $|\bar t-\bar t'|=|t-t'|$. It follows that the functions

$$a_i,\ a,\ \frac{\partial a_i}{\partial x_j},\ \frac{\partial a_i}{\partial y},\ \frac{\partial a_i}{\partial z_j}, \qquad i,j\in\{1,\dots,d\},$$

fulfill the same Hölder and Lipschitz conditions as sum and product of Hölder and Lipschitz continuous functions where in the products all functions are bounded. Now, since $t\in[0,T]$, for $|y|\le K_y\in\mathbb{R}_{>0}$ and $|z|\le K_z$ it easily follows that these functions fulfill Hölder conditions in t, y, and z with exponents $\frac{\alpha}{2}$, α and α. From Assumptions 2.1(iii) and 6.1(ii) in connection with $|z|\le K_z$ it follows that that each $f\in\big\{a_i,a,\frac{\partial a_i}{\partial x_j},\frac{\partial a_i}{\partial y},\frac{\partial a_i}{\partial z_j}\mid i,j\in\{1,\dots,d\}\big\}$ can be bounded by some constant K_f uniformly in (t,x), that is for all $(t,x)\in[0,T]\times\mathbb{R}^d$ it holds that

$$|f(t,x,0,0)|\le K_f.$$

Consequently, each function $f\in\big\{a_i,a,\frac{\partial a_i}{\partial x_j},\frac{\partial a_i}{\partial y},\frac{\partial a_i}{\partial z_j}\mid i,j\in\{1,\dots,d\}\big\}$ fulfills a Hölder condition of exponent α in x:

$$|f(t,x,y,z)-f(t,x',y,z)|\le(H_b\vee 2(H_fK_yK_z+K_f))|x-x'|^\alpha$$

for all $(t,y,z)\in[0,T]\times\mathbb{R}\times(\mathbb{R}^d)^*$ and any two $x,x'\in\mathbb{R}^d$. Lastly, from Assumptions 2.1(ii)–(iii) and 6.1(ii), we obtain

$$\sum_{i=1}^d\left(|a_i(t,x,y,z)|+\left|\frac{\partial a_i(t,x,y,z)}{\partial y}\right|\right)(1+|z|)+\sum_{i=1}^d\sum_{j=1}^d\left|\frac{\partial a_i(t,x,y,z)}{\partial x_j}\right|+|a(t,x,y,z)|$$

$$\leq \sum_{i=1}^{d}\left(\frac{1}{2}d^2K_\sigma^2|z| + d^2H_\sigma K_\sigma|z|\right)(1+|z|) + \sum_{i=1}^{d}\sum_{j=1}^{d} d^2H_\sigma K_\sigma|z|$$
$$+ dK_b|z| + K_F + H_F(|y| + K_\sigma|z|) + d^3H_\sigma K_\sigma|z| + d^3H_\sigma K_\sigma|z|^2,$$

i. e., for bounded y, the LHS admits a bound of the form $\mu(1+|z|)^2$ for a suitable constant $\mu \in \mathbb{R}_{>0}$. Together with the uniform ellipticity condition (2), solvability condition (c) is fulfilled.

Now, noting that

$$\frac{\mathrm{d}}{\mathrm{d}x_i}a_i(t,x,\tilde{u}(t,x),\tilde{u}_x(t,x)) = \frac{\partial a_i(t,x,\tilde{u}(t,x),\tilde{u}_x(t,x))}{\partial x_i} + \frac{\partial a_i(t,x,\tilde{u}(t,x),\tilde{u}_x(t,x))}{\partial y}\tilde{u}_{x_i}(t,x)$$
$$+ \sum_{k=1}^{n}\frac{\partial a_i(t,x,\tilde{u}(t,x),\tilde{u}_x(t,x))}{\partial z_k}\tilde{u}_{x_i x_k}(t,x)$$

and a classical solution $\tilde{u}$ fulfills $\tilde{u}_{x_i x_k}(t,x) = \tilde{u}_{x_k x_i}(t,x)$ for all $i,k \in \{1,\ldots,d\}$, from the definition of a it follows that

$$(14) \Leftrightarrow \begin{cases} \tilde{u}_t(t,x) = \frac{1}{2}\mathrm{tr}\big[\tilde{u}_{xx}(t,x)(\sigma\sigma^\top)(\bar{t},x,\tilde{u}(t,x))\big] - F(\bar{t},x,\tilde{u}(t,x),\tilde{u}_x(t,x)\sigma(\bar{t},x,\tilde{u}(t,x))) \\ \qquad + \tilde{u}_x(t,x)b(\bar{t},x,\tilde{u}(t,x),\tilde{u}_x(t,x)\sigma(\bar{t},x,\tilde{u}(t,x))), \quad (t,x) \in [0,T]\times\mathbb{R}^d, \\ \tilde{u}(0,x) = g(x), \quad x \in \mathbb{R}^d, \end{cases}$$
$$\Leftrightarrow \begin{cases} \tilde{u}_t(\bar{t},x) = F(t,x,\tilde{u}(\bar{t},x),\tilde{u}_x(\bar{t},x)\sigma(t,x,\tilde{u}(\bar{t},x))) - \frac{1}{2}\mathrm{tr}\big[\tilde{u}_{xx}(\bar{t},x)(\sigma\sigma^\top)(t,x,\tilde{u}(\bar{t},x))\big] \\ \qquad - \tilde{u}_x(\bar{t},x)b(t,x,\tilde{u}(\bar{t},x),\tilde{u}_x(\bar{t},x)\sigma(t,x,\tilde{u}(\bar{t},x))), \quad (t,x) \in [0,T]\times\mathbb{R}^d, \\ \tilde{u}(\bar{T},x) = g(x), \quad x \in \mathbb{R}^d, \end{cases}$$

Consequently, we can define a desired solution u of the parabolic PDE (3) via

$$u\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}, \quad u(t,x) := \tilde{u}(\bar{t},x). \qquad \square$$

References

[1] K. Andersson, A. Andersson, and C. W. Oosterlee. "Convergence of a robust deep FBSDE method for stochastic control". In: *SIAM J. Sci. Comput.* 45.1 (2023), A226–A255. DOI: 10.1137/22M1478057.

[2] F. Antonelli. "Backward-forward stochastic differential equations". In: *Ann. Appl. Probab.* 3.3 (1993), pp. 777–793.

[3] C. Bender and J. Steiner. "A posteriori estimates for backward SDEs". In: *SIAM/ASA J. Uncertain. Quantif.* 1.1 (2013), pp. 139–163. DOI: 10.1137/120878689.

[4] C. Bender and J. Zhang. "Time discretization and Markovian iteration for coupled FBSDEs". In: *Ann. Appl. Probab.* 18.1 (2008), pp. 143–177. DOI: 10.1214/07-AAP448.

[5] A. Bensoussan. "Lectures on stochastic control". In: *Nonlinear filtering and stochastic control (Cortona, 1981)*. Vol. 972. Lecture Notes in Math. Springer, Berlin-New York, 1982, pp. 1–62.

[6] J.-M. Bismut. “An introductory approach to duality in optimal stochastic control”. In: *SIAM Rev.* 20.1 (1978), pp. 62–78. DOI: 10.1137/1020004.

[7] J. Chessari, R. Kawai, Y. Shinozaki, and T. Yamada. “Numerical methods for backward stochastic differential equations: a survey”. In: *Probab. Surv.* 20 (2023), pp. 486–567. DOI: 10.1214/23-ps18.

[8] J. Cvitanić and J. Zhang. “The steepest descent method for forward-backward SDEs”. In: *Electron. J. Probab.* 10 (2005), pp. 1468–1495. DOI: 10.1214/EJP.v10-295.

[9] F. Delarue and S. Menozzi. “A forward–backward stochastic algorithm for quasi-linear PDEs”. In: *Ann. Appl. Probab.* 16.1 (2006), pp. 140–184. DOI: 10.1214/105051605000000674.

[10] J. Douglas Jr., J. Ma, and P. Protter. “Numerical methods for forward-backward stochastic differential equations”. In: *Ann. Appl. Probab.* 6.3 (1996), pp. 940–968. DOI: 10.1214/aoap/1034968235.

[11] W. E, J. Han, and A. Jentzen. “Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations”. In: *Commun. Math. Stat.* 5.4 (2017), pp. 349–380. DOI: 10.1007/s40304-017-0117-6.

[12] C. Gao, S. Gao, R. Hu, and Z. Zhu. “Convergence of the backward deep BSDE method with applications to optimal stopping problems”. In: *SIAM J. Financial Math.* 14.4 (2023), pp. 1290–1303. DOI: 10.1137/22M1539952.

[13] J. Han, A. Jentzen, and W. E. “Solving high-dimensional partial differential equations using deep learning”. In: *Proc. Natl. Acad. Sci. USA* 115.34 (2018), pp. 8505–8510. DOI: 10.1073/pnas.1718942115.

[14] J. Han and J. Long. “Convergence of the deep BSDE method for coupled FBSDEs”. In: *Probab. Uncertain. Quant. Risk* 5.5 (2020), pp. 1–33. DOI: 10.1186/s41546-020-00047-w.

[15] Z. Huang, B. Negyesi, and C. W. Oosterlee. “Convergence of the deep BSDE method for stochastic control problems formulated through the stochastic maximum principle”. In: *Math. Comput. Simulation* 227 (2025), pp. 553–568. DOI: 10.1016/j.matcom.2024.08.002.

[16] Y. Jiang and J. Li. “Convergence of the deep BSDE method for FBSDEs with non-Lipschitz coefficients”. In: *Probab. Uncertain. Quant. Risk* 6.4 (2021), pp. 391–408. DOI: 10.3934/puqr.2021019.

[17] D. P. Kingma and J. Ba. *Adam: a method for stochastic optimization.* 2017. arXiv: 1412.6980v9 [cs.LG].

[18] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations.* Springer, 1992. DOI: 10.1007/978-3-662-12616-5.

[19] O. A. Ladyženskaja, V. A. Solonnikov, and Ural’ceva. *Linear and quasilinear equations of parabolic type.* American Mathematical Society, 1968.

[20] J. Ma, P. Protter, and J. Yong. “Solving forward-backward stochastic differential equations explicitly—a four step scheme”. In: *Probab. Theory Related Fields* 98.3 (1994), pp. 339–359. DOI: 10.1007/BF01192258.

[21] J. Ma, J. Shen, and Y. Zhao. "On numerical approximations of forward-backward stochastic differential equations". In: *SIAM J. Numer. Anal.* 46.5 (2008), pp. 2636–2661. DOI: 10.1137/06067393X.

[22] J. Ma, Z. Wu, D. Zhang, and J. Zhang. "On well-posedness of forward-backward SDEs—a unified approach". In: *Ann. Appl. Probab.* 25.4 (2015), pp. 2168–2214. DOI: 10.1214/14-AAP1046.

[23] J. Ma and J. Yong. *Forward-backward stochastic differential equations and their applications.* Springer, 2007. DOI: 10.1007/978-3-540-48831-6.

[24] X. Mao. *Stochastic differential equations and applications.* 2nd ed. Woodhead Publishing, 2008.

[25] B. Negyesi, Z. Huang, and C. W. Oosterlee. "Generalized convergence of the deep BSDE method: a step towards fully-coupled FBSDEs and applications in stochastic control". In: 11.2 (2026), pp. 223–246. DOI: 10.3934/puqr.2026010.

[26] N. Nüsken and L. Richter. "Interpolating between BSDEs and PINNs: deep learning for elliptic and parabolic boundary value problems". In: *J. Mach. Learn.* 2.1 (2023), pp. 31–64. DOI: 10.4208/jml.220416.

[27] E. Pardoux and S. Peng. "Backward stochastic differential equations and quasilinear parabolic partial differential equations". In: *Stochastic partial differential equations and their applications.* Ed. by B. L. Rozovskii and R. B. Sowers. Springer, 1992, pp. 200–217. DOI: 10.1007/BFb0007334.

[28] E. Pardoux and S. Tang. "Forward-backward stochastic differential equations and quasilinear parabolic PDEs". In: *Probab. Theory Related Fields* 114.2 (1999), pp. 123–150. DOI: 10.1007/s004409970001.

[29] M. Raissi, P. Perdikaris, and G. E. Karniadakis. "Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations". In: *J. Comput. Phys.* 378 (2019), pp. 686–707. DOI: 10.1016/j.jcp.2018.10.045.

[30] C. Reisinger, W. Stockinger, and Y. Zhang. "A posteriori error estimates for fully coupled McKean-Vlasov forward-backward SDEs". In: *IMA J. Numer. Anal.* 44.4 (2024), pp. 2323–2369. DOI: 10.1093/imanum/drad060.

[31] J. Yong and X. Y. Zhou. *Stochastic controls.* Springer, 1999. DOI: 10.1007/978-1-4612-1466-3.

Appendix A: Proof of Lemma 5.8

From the recursion of the Euler-Maruyama scheme (8), we obtain

$$\tilde{X}_{t_i}^{\eta,\pi} = \tilde{x}_0 + \sum_{j=0}^{i-1} \big[b^\rho(t_j, \tilde{X}_{t_j}^{\eta,\pi})\Delta_j + \sigma^\rho(t_j, \tilde{X}_{t_j}^{\eta,\pi})\Delta W_j \big].$$

It directly follows that, for all $t \in [0,T]$, we can express $\tilde{X}_t^{\eta,\pi}$ as

$$\tilde{X}_t^{\eta,\pi} = \tilde{x}_0 + \int_0^{\Pi(t)} b^\rho(\Pi(s), \tilde{X}_s^{\eta,\pi})\,\mathrm{d}s + \int_0^{\Pi(t)} \sigma^\rho(\Pi(s), \tilde{X}_s^{\eta,\pi})\,\mathrm{d}W_s. \tag{15}$$

Let $t \in [0,T]$. Inserting the definition of $\tilde{X}_t^\eta$ and Equation (15) for $\tilde{X}_t^{\eta,\pi}$, we obtain

$$\begin{aligned}
&\mathbb{E}\big[|\tilde{X}_t^\eta - \tilde{X}_t^{\eta,\pi}|^2\big] \\
&\leq 4\Bigg(\mathbb{E}\Bigg[\Bigg|\int_0^{\Pi(t)} b^\rho(s,\tilde{X}_s^\eta) - b^\rho(\Pi(s),\tilde{X}_s^{\eta,\pi})\,\mathrm{d}s\Bigg|^2\Bigg] + \mathbb{E}\Bigg[\Bigg|\int_{\Pi(t)}^t b^\rho(s,\tilde{X}_s^\eta)\,\mathrm{d}s\Bigg|^2\Bigg] \\
&\quad + \mathbb{E}\Bigg[\Bigg|\int_0^{\Pi(t)} \sigma^\rho(s,\tilde{X}_s^\eta) - \sigma^\rho(\Pi(s),\tilde{X}_s^{\eta,\pi})\,\mathrm{d}W_s\Bigg|^2\Bigg] + \mathbb{E}\Bigg[\Bigg|\int_{\Pi(t)}^t \sigma^\rho(s,\tilde{X}_s^\eta)\,\mathrm{d}W_s\Bigg|^2\Bigg]\Bigg) \\
&=: 4\big((\mathrm{I}) + (\mathrm{II}) + (\mathrm{III}) + (\mathrm{IV})\big).
\end{aligned}$$

Using Hölder's inequality, the $\frac{1}{2}$-Hölder and Lipschitz continuity of b^ρ and Fubini's theorem, the first term can be bounded by

$$\begin{aligned}
(\mathrm{I}) &\leq \Pi(t)\mathbb{E}\Bigg[\int_0^{\Pi(t)} |b^\rho(s,\tilde{X}_s^\eta) - b^\rho(\Pi(s),\tilde{X}_s^{\eta,\pi})|^2\,\mathrm{d}s\Bigg] \\
&\leq TH_{b^\rho}^2\mathbb{E}\Bigg[\int_0^{\Pi(t)} \big(|s-\Pi(s)|^{1/2} + |\tilde{X}_s^\eta - \tilde{X}_s^{\eta,\pi}|\big)^2\,\mathrm{d}s\Bigg] \\
&\leq 2TH_{b^\rho}^2\Bigg(\mathbb{E}\Bigg[\int_0^{\Pi(t)} |s-\Pi(s)|\,\mathrm{d}s\Bigg] + \mathbb{E}\Bigg[\int_0^{\Pi(t)} |\tilde{X}_s^\eta - \tilde{X}_s^{\eta,\pi}|^2\,\mathrm{d}s\Bigg]\Bigg) \\
&\leq 2TH_{b^\rho}^2\Bigg(T|\pi| + \int_0^t \mathbb{E}\big[|\tilde{X}_s^\eta - \tilde{X}_s^{\eta,\pi}|^2\big]\,\mathrm{d}s\Bigg).
\end{aligned}$$

For (III), using Itô's isometry, the Hölder and Lipschitz condition on σ^ρ and Fubini's theorem, we obtain

$$\begin{aligned}
(\mathrm{III}) &= \mathbb{E}\Bigg[\int_0^{\Pi(t)} |\sigma^\rho(s,\tilde{X}_s^\eta) - \sigma^\rho(\Pi(s),\tilde{X}_s^{\eta,\pi})|^2\,\mathrm{d}s\Bigg] \\
&\leq H_{\sigma^\rho}^2\mathbb{E}\Bigg[\int_0^{\Pi(t)} \big(|s-\Pi(s)|^{1/2} + |\tilde{X}_s^\eta - \tilde{X}_s^{\eta,\pi}|\big)^2\,\mathrm{d}s\Bigg] \\
&\leq 2H_{\sigma^\rho}^2\mathbb{E}\Bigg[\int_0^{\Pi(t)} |s-\Pi(s)| + |\tilde{X}_s^\eta - \tilde{X}_s^{\eta,\pi}|^2\,\mathrm{d}s\Bigg] \\
&\leq 2H_{\sigma^\rho}^2\Bigg(T|\pi| + \int_0^t \mathbb{E}\big[|\tilde{X}_s^\eta - \tilde{X}_s^{\eta,\pi}|^2\big]\,\mathrm{d}s\Bigg).
\end{aligned}$$

Applying Hölder's inequality and noticing that b^ρ fulfills the linear growth condition of Lemma 5.4, we see that the second term fulfills

$$(\mathrm{II}) \leq |\pi|L_{b^\rho}^2\mathbb{E}\Bigg[\int_{\Pi(t)}^t 1 + |\tilde{X}_s^\eta|^2\,\mathrm{d}s\Bigg] \leq |\pi|^2L_{b^\rho}^2(1+c_6),$$

where the last inequality follows from Lemma 5.6. Similarly, now using Itô's isometry and boundedness of σ^ρ from Lemma 5.4, for (IV) we obtain

$$(\mathrm{IV}) \leq \mathbb{E}\Bigg[\int_{\Pi(t)}^t K_{\sigma^\rho}^2\,\mathrm{d}s\Bigg] \leq K_{\sigma^\rho}^2|\pi|.$$

Putting all bounds together yields

$$\mathbb{E}\big[|\tilde{X}_t^\eta - \tilde{X}_t^{\eta,\pi}|^2\big] \le 8(TH_{b^\rho}^2 + H_{\sigma^\rho}^2)\int_0^t \mathbb{E}\big[|\tilde{X}_s^\eta - \tilde{X}_s^{\eta,\pi}|^2\big]\,\mathrm{d}s + 4\big(2T(TH_{b^\rho}^2 + H_{\sigma^\rho}^2) + L_{b^\rho}^2|\pi|(1+c_6) + K_{\sigma^\rho}^2\big)|\pi|.$$

Applying Gronwall's lemma finishes the proof.

Appendix B: Upper constant bounds depending on K_ρ and H_ρ

	$C(1+K_\rho^p)$	$C(1+H_\rho)$	$C(1+K_\rho+H_\rho)$	$Ce^{CK_\rho^p}$	$Ce^{C(K_\rho^2+H_\rho^2)}$
L_{b^ρ}	$\times_{p=1}$				
K_{σ^ρ}	$\times_{p=1}$				
H_{b^ρ}		$\times$			
H_{σ^ρ}		$\times$			
$H_\zeta^{(t,x)}$	$\times_{p=1}$				
$H_\zeta^{(y,z)}$	$\times_{p=1}$				
H_1	$\times_{p=1}$				
H_2			$\times$		
α_π	$\times_{p=2}$				
β_π	$\times_{p=2}$				
Γ	$\times_{p=4}$				
d_1				$\times_{p=4}$	
d_2				$\times_{p=2}$	
d_3				$\times_{p=2}$	
d_4				$\times_{p=2}$	
d_5				$\times_{p=2}$	
$\tilde{d}_2$				$\times_{p=2}$	
$\tilde{d}_3$				$\times_{p=2}$	
c_1				$\times_{p=4}$	
c_2				$\times_{p=4}$	
c_3					$\times$
c_4					$\times$
c_5					$\times$
c_6				$\times_{p=2}$	
c_7				$\times_{p=4}$	
c_8				$\times_{p=4}$	
c_9					$\times$
$\tilde{c}_4$					$\times$
$\tilde{c}_5$					$\times$
$\tilde{c}_9$					$\times$

For all other constants, including H_3, c_{10} and c_{11}, there exist upper bounds that do not depend on K_ρ or H_ρ.