CHAOTIC HEDGING WITH ITERATED INTEGRALS AND NEURAL NETWORKS

ARIEL NEUFELD AND PHILIPP SCHMOCKER

ABSTRACT. In this paper, we derive an L^p-chaos expansion based on iterated Stratonovich integrals with respect to a given exponentially integrable continuous semimartingale. By omitting the orthogonality of the expansion, we show that every p-integrable functional, $p \in [1,\infty)$, can be approximated by a finite sum of iterated Stratonovich integrals. Using (possibly random) neural networks as integrands, we therefere obtain universal approximation results for p-integrable financial derivatives in the L^p-sense. Moreover, we can approximately solve the L^p-hedging problem (coinciding for $p=2$ with the quadratic hedging problem), where the approximating hedging strategy can be computed in closed form within short runtime.

1. INTRODUCTION

We address the problem of pricing and hedging a financial derivative in a market with finite time horizon $T > 0$ that consists of $d \in \mathbb{N}$ risky assets. The price processes of these risky assets are modelled by a d-dimensional continuous semimartingale $X := (X_t)_{t\in[0,T]}$, which has a decomposition $X_t = X_0 + A_t + M_t$ into an adapted process $A := (A_t)_{t\in[0,T]}$ of finite variation and a local martingale $M := (M_t)_{t\in[0,T]}$. In this paper, we assume that X is exponentially integrable and that A as well as $\langle M \rangle$ have moments of all orders (see Assumption 2.1), which includes in particular affine and some polynomial diffusions as well as some stochastic volatility models (see [8, 17, 21] and also Section 2+6).

Our chaotic hedging approach relies on the chaos expansion of a financial derivative, which can be understood as Taylor formula for random variables. In the case of X being a Brownian motion, the Wiener-Ito chaos decomposition (see [32, Theorem 4.2] and also [3, 63]) ensures that every square-integrable functional of the Brownian motion can be represented as infinite sum of orthogonal multiple integrals, which can be rewritten as iterated integrals with respect to that Brownian motion (see, e.g., [15, p. 10]).

In our paper, we omit the orthogonality of the expansion and derive an L^p-chaos expansion based on iterated Stratonovich integrals with respect to the given exponentially integrable continuous semimartingale X. More precisely, we show that every financial derivative $G \in L^p(\Omega, \mathcal{F}_T, \mathbb{P})$, $p \in [1,\infty)$, can be approximated by iterated Stratonovich integrals (see Theorem 4.2), i.e. for every $\varepsilon > 0$ there exists some $N \in \mathbb{N}$ and deterministic functions $g_n : [0,T]^n \to (\mathbb{R}^d)^{\otimes n}$, $n = 0, ..., N$, such that[1]

$$\mathbb{E}\left[\left|G - \sum_{n=0}^{N} J_n^\circ(g_n)\right|^p\right]^{\frac{1}{p}} < \varepsilon, \tag{1}$$

where $J_n^\circ(g_n)_T$ denotes the n-fold iterated Stratonovich integral of $g_n : [0,T]^n \to (\mathbb{R}^d)^{\otimes n}$, defined by

$$J_n^\circ(g_n)_T := \sum_{j=1}^{m_n} \int_0^T \left(\int_0^{t_{n-1}} \cdots \left(\int_0^{t_2} \circ \underbrace{dW(g_{n,j,1})_{t_1}}_{=g_{n,j,1}(t_1)^\top dX_{t_1}} \right) \cdots \circ \underbrace{dW(g_{n,j,n-1})_{t_{n-1}}}_{=g_{n,j,n-1}(t_{n-1})^\top dX_{t_{n-1}}} \right) \circ \underbrace{dW(g_{n,j,n})_{t_n}}_{=g_{n,j,n}(t_n)^\top dX_{t_n}}$$

for $g_n : [0,T]^n \to (\mathbb{R}^d)^{\otimes n}$ being of the form $g_n = \sum_{j=1}^{m_n} g_{n,j,1} \otimes \cdots \otimes g_{n,j,n}$ for some $m_n \in \mathbb{N}$ and functions $g_{n,j,1}, ..., g_{n,j,n} : [0,T] \to \mathbb{R}^d$ that are regular enough such that the Stratonovich integral[2] is well defined at each iteration. Hereby, $W(g_{n,j,k})_t := \int_0^t g_{n,j,k}(s)^\top dX_s$ denotes the stochastic Ito integral of $g_{n,j,k} : [0,T] \to \mathbb{R}^d$ with respect to X up to time $t \in [0,T]$, for $j = 1, ..., m$ and $k = 1, ..., n$.

Date: June 6, 2025.

Key words and phrases. Chaos expansion, iterated Stratonovich integrals, machine learning, neural networks, random neural networks, universal approximation, hedging, quadratic hedging.

Financial support by the Nanyang Assistant Professorship Grant (NAP Grant) *Machine Learning based Algorithms in Finance and Insurance* is gratefully acknowledged.

[1] Hereby, we denote by $\mathbb{F} := (\mathcal{F}_t)_{t\in[0,T]}$ the usual $\mathbb{P}$-augmented filtration generated by X. Moreover, $(\mathbb{R}^d)^{\otimes n} := \mathbb{R}^d \otimes \cdots \otimes \mathbb{R}^d$ represents the tensor product of $\mathbb{R}^d$, with convention $(\mathbb{R}^d)^{\otimes 0} := \mathbb{R}$.

[2] For a sufficiently integrable $\mathbb{F}$-predictable process $\theta := (\theta_t)_{t\in[0,T]}$ and an $\mathbb{F}$-adapted process $Y := (Y_t)_{t\in[0,T]}$, the Stratonovich integral $\int_0^t \theta_s \circ dY_s$ is defined as the $L^2(\mathbb{P})$-limit of $\sum_{k=1}^n \frac{1}{2}(\theta_{t_k} + \theta_{t_{k+1}})(Y_{t_{k+1}} - Y_{t_k})$ over partitions of the form $0 \le t_0 < t_1 < ... < t_n \le T$, as the mesh $\max_{k=1,...,n} |t_k - t_{k-1}|$ goes to zero (see, e.g., [54, Chapter V.5]).

In this case, (1) can be understood as *chaos expansion* of $L^p(\Omega, \mathcal{F}_T, \mathbb{P})$ with respect to X. This extends classical chaos expansions in the literature (proven for the compensated Poisson process [33], Azéma martingales [20], some Lévy processes via Teugel martingales [50], and other specific martingales in [16, 34]) from orthogonal decompositions of $L^2(\Omega, \mathcal{F}_T, \mathbb{P})$ using iterated *Ito integrals* to denseness of iterated *Stratonovich integrals* in $L^p(\Omega, \mathcal{F}_T, \mathbb{P})$, where $p \in [1, \infty)$. Our approach exploits the connection between iterated Stratonovich integrals and monomials of X (see Proposition 3.10), yielding denseness in $L^p(\Omega, \mathcal{F}_T, \mathbb{P})$ at the cost of using the less natural Stratonovich integral instead of the Ito integral.

As an application, we also consider the following L^p-hedging problem. For a given financial derivative $G \in L^p(\Omega, \mathcal{F}_T, \mathbb{P})$, we aim to find an optimal initial endowment $c \in \mathbb{R}$ and an optimal trading strategy $\theta := (\theta_t)_{t\in[0,T]}$ such that the L^p-hedging error $\left\| G - c - \int_0^T \theta_t^\top dX_t \right\|_{L^p(\mathbb{P})}$ is minimized, which coincides for $p = 2$ with the quadratic hedging approaches in [52, 61]. By using the chaos expansion in (1) together with a denseness result for $\mathbb{F}$-predictable processes, we can approximately solve the L^p-hedging problem (see Theorem 4.4), i.e. for every $\varepsilon > 0$ there exist some $g_0 \in \mathbb{R}$ and $N, m_n \in \mathbb{N}$ as well as functions $g_{n,j,0}, g_{n,j,1} : [0,T] \to \mathbb{R}^d$, $n = 1, ..., N$ and $j = 1, ..., m_n$, such that

$$\left\| G - g_0 - \int_0^T \left(\vartheta_t^{g_{1:N}}\right)^\top dX_t \right\|_{L^p(\mathbb{P})} \leq \varepsilon + \inf_{(c,\theta)\in\mathbb{R}\times\Theta^p(X)} \left\| G - c - \int_0^T \theta_t^\top dX_t \right\|_{L^p(\mathbb{P})} \tag{2}$$

with $\vartheta_t^{g_{1:N}} := \sum_{n=1}^N \sum_{j=1}^{m_n} \frac{W(g_{n,j,0})_t^{n-1}}{(n-1)!} g_{n,j,1}(t)$, where $\Theta^p(X)$ denotes the space of $\mathbb{F}$-predictable $\mathbb{R}^d$-valued processes satisfying some integrability conditions (see (10)). We highlight that (2), unlike the chaos expansion in (1), employs classical Ito integrals as typically used in the financial context.

The machine learning application of this paper then consists of using neural networks to learn the integrands $g_n : [0,T]^n \to (\mathbb{R}^d)^{\otimes n}$ in the chaos expansion (1) as well as the functions $g_{n,j,0}, g_{n,j,1} : [0,T] \to \mathbb{R}^d$ in the approximation of the L^p-hedging problem (2). We theoretically justify these neural network applications by lifting the universal approximation property of neural networks to a universal approximation result for financial derivatives (see Theorem 5.4). In addition, we extend this universal approximation result to random neural networks with randomly initialized weights and biases (see Theorem 5.9), which are inspired by the works on extreme learning machines and random feature learning (see, e.g., [23, 24, 26, 30, 55]). In this case, only the linear readouts needs to be trained, which can be efficiently performed, e.g., by the least squares method. Furthermore, we show that the L^p-hedging problem described above can also be approximately solved using (possibly random) neural networks.

Altogether, our chaotic hedging approach contributes to already existing applications of chaos expansions, e.g., for option pricing (see [39, 40]) and for solving SDEs (see [64]), as well as to other successful neural networks applications in mathematical finance (see, e.g., [2, 9, 19, 28, 47, 48, 58, 60, 62]). Moreover, our results can be related to the universal approximation results for signatures that satisfy a similar universality as neural networks (see [5, 10, 11, 37, 44] and also Remark 5.10).

1.1. **Outline.** In Section 2, we introduce the main setting including exponentially integrable continuous semimartingale. In Section 3, we define iterated Stratonovich integrals. In Section 4, we show a chaos expansion using Stratonovich integrals and approximately solve the L^p-hedging problem. In Section 5, we derive universal approximation results using (possibly random) neural networks, which are then applied in Section 6 to numerically solve the L^2-hedging problem. Finally, Section 7 contains all proofs.

1.2. **Notation.** As usual, we denote by $\mathbb{N} = \{1, 2, ...\}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ the sets of natural numbers, while $\mathbb{R}$ and $\mathbb{C}$ represent the real and complex numbers (with complex unit $\mathbf{i} := \sqrt{-1}$), respectively, where $s \wedge t := \min(s,t)$, for $s, t \in \mathbb{R}$. Moreover, for $d \in \mathbb{N}$, we denote by $\mathbb{R}^d$ the Euclidean space equipped with the norm $\|x\| = \left(\sum_{i=1}^d x_i^2\right)^{1/2}$. In addition, for $d, l \in \mathbb{N}$, we denote by $\mathbb{R}^{d\times l}$ the vector space of matrices $A = (a_{i,j})_{i=1,...,d,\, j=1,...,l}$ equipped with the Frobenius norm $\|A\|_F = \left(\sum_{i=1}^d \sum_{j=1}^l a_{i,j}^2\right)^{1/2}$. For $d = l$, we denote by $\mathbb{S}_+^d \subset \mathbb{R}^{d\times d}$ the cone of symmetric non-negative definite matrices, use the notation $I_d \in \mathbb{S}_+^d$ for the identity matrix, and define $\mathrm{tr}(A) := \sum_{i=1}^d a_{i,i}$ for $A = (a_{i,j})_{i,j=1,...,d} \in \mathbb{R}^{d\times d}$.

Furthermore, for $k, d \in \mathbb{N}$ and $E \subseteq \mathbb{R}^k$, we denote by $C(E; \mathbb{R}^d)$ the Banach space of continuous functions $f : E \to \mathbb{R}^d$ equipped with the supremum norm $\|f\|_\infty := \sup_{t\in E} \|f(t)\|$, where we abbreviate $C(E) := C(E; \mathbb{R})$. In addition, for $p \in [1, \infty)$ and a measure space (S, Σ, μ), we denote by $L^p(S, \Sigma, \mu; \mathbb{R}^d)$ the Banach space of (equivalence classes of) $\Sigma/\mathcal{B}(\mathbb{R}^d)$-measurable functions $f : S \to \mathbb{R}^d$ with finite norm $\|f\|_{L^p(\mu)} := \left(\int_S \|f(t)\|^p \mu(dt)\right)^{1/p}$, where we abbreviate $L^p(S, \Sigma, \mu) := L^p(S, \Sigma, \mu; \mathbb{R})$. Moreover, if S is a topological space, we denote by $\mathcal{B}(S)$ the Borel σ-algebra of S.

2. EXPONENTIALLY INTEGRABLE SEMIMARTINGALES

Throughout this paper, we fix a finite time horizon $T > 0$ and a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. For $d \in \mathbb{N}$, the financial market is modelled by a d-dimensional stochastic process $X := \left(X_t^1, ..., X_t^d\right)^\top_{t\in[0,T]}$ that is assumed to be a continuous semimartingale with respect to the usual $\mathbb{P}$-augmented filtration[3] generated by X. This means that there exists a continuous $\mathbb{F}$-adapted process $A := (A_t)_{t\in[0,T]}$ of finite variation with $A_0 = 0$ and a continuous local martingale $M := (M_t)_{t\in[0,T]}$ with $M_0 = 0$ such that for every $t \in [0,T]$ it holds that $X_t = X_0 + A_t + M_t$, where we assume that the initial value $X_0 \in \mathbb{R}^d$ is $\mathbb{P}$-a.s. constant. Moreover, we denote by $\langle X \rangle := (\langle X^i, X^j \rangle_t)_{t\in[0,T],\, i,j=1,...,d}$ the $\mathbb{S}_+^d$-valued quadratic variation of X satisfying $\langle X \rangle = \langle M \rangle$, whereas $\int_0^T \Vert dA_t \Vert$ represents the total variation of A over $[0,T]$.

Assumption 2.1. *Let X be a continuous semimartingale with decomposition $X_t = X_0 + A_t + M_t$, $t \in [0,T]$, into a $\mathbb{P}$-a.s. constant initial value $X_0 \in \mathbb{R}^d$, a continuous $\mathbb{F}$-adapted process $A = (A_t)_{t\in[0,T]}$ of finite variation with $A_0 = 0$, and a continuous local martingale $M = (M_t)_{t\in[0,T]}$ with $M_0 = 0$. Moreover, assume the following integrability conditions:*

(i) *There exists $\varepsilon > 0$ such that for every $t \in [0,T]$ it holds that $\mathbb{E}\left[\exp\left(\varepsilon \Vert X_t \Vert\right)\right] < \infty$.*
(ii) *For every $\gamma \in [1,\infty)$ it holds that $\mathbb{E}\left[\left(\int_0^T \Vert dA_t \Vert\right)^\gamma\right] < \infty$.*
(iii) *For every $\gamma \in [1,\infty)$ it holds that $\mathbb{E}\left[\Vert \langle M \rangle_T \Vert_F^\gamma\right] < \infty$.*

Remark 2.2. *Note that $\mathbb{E}\left[\exp\left(\varepsilon \int_0^T \Vert dA_t \Vert\right)\right] < \infty$ and $\mathbb{E}\left[\exp\left(\varepsilon \Vert \langle M \rangle_T \Vert_F\right)\right] < \infty$ for some $\varepsilon > 0$ are sufficient conditions for (ii) and (iii), respectively.*

Let us give some examples of continuous semimartingales satisfying Assumption 2.1, which includes affine diffusions and some polynomial diffusions (see [8, 17, 21]). To this end, we denote by $\mathrm{Pol}_n(\mathbb{R}^d)$ the set of polynomial functions $\mathbb{R}^d \ni x := (x_1, ..., x_d) \mapsto \sum_{\alpha\in\mathbb{N}_0^d,\, |\alpha|\leq n} c_\alpha x_1^{\alpha_1} \cdots x_d^{\alpha_d} \in \mathbb{R}$, with some coefficients $c_\alpha \in \mathbb{R}$, where $|\alpha| = \alpha_1 + ... + \alpha_n$ for $\alpha := (\alpha_1, ..., \alpha_d) \in \mathbb{N}_0^d$.

Example 2.3 (Polynomial diffusions)**.** *For a subset $E \subseteq \mathbb{R}^d$ and some functions $a := (a_{i,j})_{i,j=1,...,d} : E \to \mathbb{S}_+^d$ and $b := (b_1, ..., b_d)^\top : E \to \mathbb{R}^d$ with $a_{i,j} \in \mathrm{Pol}_2(\mathbb{R}^d)$ and $b_i \in \mathrm{Pol}_1(\mathbb{R}^d)$ for all $i,j = 1, ..., d$, let here $X = (X_t)_{t\in[0,T]}$ be an E-valued strong solution of the stochastic differential equation (SDE)*

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \qquad t \in [0,T], \tag{3}$$

with initial value $X_0 \in E$, where $\sigma : E \to \mathbb{R}^{d\times d}$ satisfies $\sigma(x)\sigma(x)^\top = a(x)$ for all $x \in E$, and where $B := (B_t)_{t\in[0,T]}$ is a d-dimensional Brownian motion. Then, X is a continuous semimartingale (with finite variation part $t \mapsto A_t := \int_0^t b(X_s)ds$ and local martingale part $t \mapsto M_t := \int_0^t \sigma(X_s)dB_s$), called a polynomial diffusion in the sense of [21, Definition 2.1]. Moreover, by assuming that the diffusion coefficient $a : E \to \mathbb{S}_+^d$ satisfies a linear growth condition, i.e. there exists $C_1 > 0$ such that for every $x \in E$ it holds that

$$\Vert a(x) \Vert_F \leq C_1 \left(1 + \Vert x \Vert\right), \tag{4}$$

then one can show that X satisfies Assumption 2.1. The proof can be found in Section 7.1.

For $d = 1$, we observe that $X = (X_t)_{t\in[0,T]}$ is an E-valued strong solution of the SDE

$$dX_t = (\beta_0 + \beta_1 X_t)dt + \sqrt{\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2}\, dB_t, \qquad t \in [0,T],$$

with initial value $X_0 \in E$, where $\beta_0, \beta_1, \alpha_0, \alpha_1, \alpha_2 \in \mathbb{R}$ are such that $\sqrt{\alpha_0 + \alpha_1 x + \alpha_2 x^2} \geq 0$ for all $x \in E$. Hence, if the drift vanishes (i.e. $\beta_0 = \beta_1 = 0$), there are in particular three relevant cases:

$$E = \mathbb{R} \qquad \text{with } \alpha_0 > 0,\ \alpha_1 = 0 \text{ and } \alpha_2 \geq 0, \qquad \text{e.g., Brownian motion (BM),} \tag{5a}$$

$$E = [0,\infty) \qquad \text{with } \alpha_0 = 0 \text{ and } \alpha_1, \alpha_2 \geq 0, \qquad \text{e.g., geometric BM (GBM),} \tag{5b}$$

$$E = [0,1] \qquad \text{with } \alpha_0 = 0 \text{ and } \alpha_1 = -\alpha_2, \qquad \text{e.g., Jacobi process.} \tag{5c}$$

If $\alpha_2 = 0$ or E is compact (e.g., (5c)), then the diffusion coefficient $E \ni x \mapsto a(x) := \alpha_0 + \alpha_1 x + \alpha_2 x^2 \in [0,\infty)$ satisfies the linear growth condition in (4) and X thus satisfies Assumption 2.1.

Moreover, we can also consider financial market models X with stochastic volatility such that X satisfies Assumption 2.1 (see also the numerical experiments in Section 6).

[3] The usual $\mathbb{P}$-augmented filtration generated by X is defined as the smallest filtration $\mathbb{F} := (\mathcal{F}_t)_{t\in[0,T]}$ such that $\mathbb{F}$ is complete with respect to $(\Omega, \mathcal{A}, \mathbb{P})$ and right-continuous on $[0,T)$, and X is $\mathbb{F}$-adapted (see [56, p. 45]).

3. Stochastic integral and iterated Stratonovich integrals

In this section, we introduce the stochastic integral of a deterministic function with respect to the given continuous semimartingale $X := (X_t)_{t\in[0,T]}$ having semimartingale decomposition $X_t = X_0 + A_t + M_t$, $t \in [0,T]$. Subsequently, we introduce iterated Stratonovich integrals that are defined by iteration.

3.1. **Stochastic integral.** In order to define the stochastic integral of a deterministic function, we fix some $p \in [1,\infty)$. Then, we denote by $L^p(X)$ the vector space of (equivalence classes of) càglàd (i.e., left continuous on $(0,T]$ with right limits on $[0,T)$) functions $g : [0,T] \to \mathbb{R}^d$ such that

$$\|g\|_{L^p(X)} := \mathbb{E}\left[\left(\int_0^T \left|g(t)^\top dA_t\right|\right)^p\right]^{\frac{1}{p}} + \mathbb{E}\left[\left(\int_0^T g(t)^\top d\langle M\rangle_t g(t)\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} < \infty. \tag{6}$$

Remark 3.1. *Since for every $g \in L^p(X)$ the function $t \mapsto g(T-t)$ is càdlàg (right continuous on $[0,T)$ with left limits on $(0,T]$), we conclude from [1, p. 122] that $g : [0,T] \to \mathbb{R}^d$ is $\mathcal{B}([0,T])/\mathcal{B}(\mathbb{R}^d)$-measurable and bounded, i.e. $\|g\|_\infty := \sup_{t\in[0,T]} \|g(t)\| < \infty$.*

We show that $L^p(X)$ is a normed vector space under the norm $\|\cdot\|_{L^p(X)}$ and obtain an estimate for $L^p(X)$-functions. The proofs of the results in this section can be found in Section 7.2.1.

Lemma 3.2. *Let X be a continuous semimartingale and let $p \in [1,\infty)$. Then, the following holds true:*
- (i) *$(L^p(X), \|\cdot\|_{L^p(X)})$ is a normed vector space.*
- (ii) *If Assumption 2.1 additionally holds, then there exists a constant $C_{p,T,X} > 0$ such that for every $g \in L^p(X)$ it holds that $\|g\|_{L^p(X)} \leq C_{p,T,X}\|g\|_\infty < \infty$.*

Now, we are able to introduce the stochastic integral as operator, which is defined on this $L^p(X)$-space.

Definition 3.3. *Let X be a continuous semimartingale and let $p \in [1,\infty)$. Then, for every $t \in [0,T]$, we introduce the* stochastic integral (up to time t) *as operator $W(\cdot)_t : L^p(X) \to L^p(\Omega, \mathcal{F}_t, \mathbb{P})$ defined by*

$$L^p(X) \ni g \quad \mapsto \quad W(g)_t := \int_0^t g(s)^\top dX_s \in L^p(\Omega, \mathcal{F}_t, \mathbb{P}).$$

Remark 3.4. *For every fixed $g \in L^p(X)$ and $t \in [0,T]$, we first observe that $W(g)_t = W(\mathbb{1}_{[0,t]}g)_T$. Moreover, we show in Lemma 3.5 below that the operator $W(\cdot)_t : L^p(X) \to L^p(\Omega,\mathcal{F}_t,\mathbb{P})$ is well-defined and bounded. In addition, $W(g)_t = \int_0^t g(s)^\top dX_s$ is well-defined as stochastic integral since the integrand $s \mapsto g(s)$ is left-continuous and $\mathbb{F}$-adapted, thus $\mathbb{F}$-predictable and locally bounded.*

Lemma 3.5. *Let X be a continuous semimartingale and let $p \in [1,\infty)$. Then, there exists a constant $C_{1,p} > 0$ such that for every $g \in L^p(X)$ it holds that*

$$\mathbb{E}\left[\sup_{t\in[0,T]} |W(g)_t|^p\right]^{\frac{1}{p}} \leq C_{1,p}\|g\|_{L^p(X)}.$$

Hereby, $C_{1,p} := \max(1, C_p) > 0$ consists of the constant $C_p > 0$ appearing in the classic upper Burkholder-Davis-Gundy inequality with exponent $p \in [1,\infty)$.

3.2. **Iterated Stratonovich integrals.** Let us first briefly revisit the notion of the tensor product (see [59, Chapter 1]). For any $n \in \mathbb{N}_0$, we define $(\mathbb{R}^d)^{\otimes n} := \mathbb{R}^d \otimes \cdots \otimes \mathbb{R}^d$, with convention $(\mathbb{R}^d)^{\otimes 0} \cong \mathbb{R}$. Moreover, for any $p \in [1,\infty)$, we define $L^{np}(X)^{\otimes n} := L^{np}(X) \otimes \cdots \otimes L^{np}(X)$, with convention $L^{np}(X)^{\otimes n} := \mathbb{R}$ for $n = 0$. Then, every tensor $g \in L^{np}(X)^{\otimes n}$ can be seen as linear functional acting on multilinear forms $L^{np}(X) \times \cdots \times L^{np}(X) \to \mathbb{R}$ (see [59, Section 1.1]), which can be expressed as

$$g = \sum_{j=1}^m g_{j,1} \otimes \cdots \otimes g_{j,n} \ \in \ L^{np}(X)^{\otimes n} := L^{np}(X) \otimes \cdots \otimes L^{np}(X)$$

for some $m \in \mathbb{N}$ and $g_{j,1},...,g_{j,n} \in L^{np}(X)$, $j = 1,...,m$, where the representation might not be unique (see [59, Section 1.1]). On the vector space $L^{np}(X)^{\otimes n}$, we use the projective tensor norm defined by

$$\|g\|_{L^{np}(X)^{\otimes n}} = \inf\left\{\sum_{j=1}^m \prod_{k=1}^n \|g_{j,k}\|_{L^{np}(X)} : \begin{array}{l} g = \sum_{j=1}^m g_{j,1} \otimes \cdots \otimes g_{j,n} \\ m \in \mathbb{N},\ g_{j,1},...,g_{j,n} \in L^{np}(X) \end{array}\right\}$$

and show some properties of $\|\cdot\|_{L^{np}(X)^{\otimes n}}$. The proofs of this section are given in Section 7.2.2.

Lemma 3.6. *For every $n \in \mathbb{N}$ and $p \in [1,\infty)$, $(L^{np}(X)^{\otimes n}, \|\cdot\|_{L^{np}(X)^{\otimes n}})$ is a normed vector space. Moreover, for every $g_1,...,g_n \in L^{np}(X)$, it holds that $\|g_1 \otimes \cdots \otimes g_n\|_{L^{np}(X)^{\otimes n}} = \prod_{k=1}^n \|g_k\|_{L^{np}(X)}$.*

Note that for $n = 1$, we recover $L^p(X)$ from Section 3.1, i.e. $L^p(X)^{\otimes 1} \cong L^p(X)$. Now, we can introduce iterated Stratonovich integrals with respect to X and show that they are well-defined.

Definition 3.7. *Let X be a continuous semimartingale, let $n \in \mathbb{N}$, and $p \in [1,\infty)$. Then, for every $t \in [0,T]$, we introduce the* n-fold iterated Stratonovich integral *as operator $J_n^\circ(\cdot)_t : L^{np}(X)^{\otimes n} \to L^p(\Omega,\mathcal{F}_t,\mathbb{P})$ defined by $J_0^\circ(g)_t = g$, for $g \in \mathbb{R}$, and for $n \geq 1$ by*

$$J_n^\circ(g)_t := \sum_{j=1}^m \int_0^t \left(\int_0^{t_{n-1}} \cdots \left(\int_0^{t_2} \circ \underbrace{dW(g_{j,1})_{t_1}}_{=g_{j,1}(t_1)^\top dX_{t_1}} \right) \cdots \circ \underbrace{dW(g_{j,n-1})_{t_{n-1}}}_{=g_{j,n-1}(t_{n-1})^\top dX_{t_{n-1}}} \right) \circ \underbrace{dW(g_{j,n})_{t_n}}_{=g_{j,n}(t_n)^\top dX_{t_n}},$$

for $g = \sum_{j=1}^m g_{j,1} \otimes \cdots \otimes g_{j,n} \in L^{np}(X)^{\otimes n}$.

Remark 3.8. *While we prove in Lemma 3.9 below that $J_n^\circ(\cdot) : L^{np}(X)^{\otimes n} \to L^p(\Omega,\mathcal{F}_t,\mathbb{P})$ is well-defined, linear, and bounded, we first show by induction on $n \in \mathbb{N}$ that $J_n^\circ(g_{j,1} \otimes \cdots \otimes g_{j,n})_t$ is well-defined as iterated Stratonovich integrals, for all $g_{j,1},...,g_{j,n} \in L^{np}(X)$ and $t \in [0,T]$. Indeed, for $n = 1$, we observe that $J_1^\circ(g_1)_t = \int_0^t \circ dW(g_1)_s = W(g_1)_t$ is by Remark 3.4 well-defined as stochastic integral. Now, if $J_{n-1}^\circ(g_{j,1} \otimes \cdots \otimes g_{j,n-1})_s$ is well-defined, for all $s \in [0,t]$ and some $n \in \mathbb{N} \cap [2,\infty)$, then*

$$\begin{aligned} &J_n^\circ(g_{j,1} \otimes \cdots \otimes g_{j,n})_t = \int_0^t J_{n-1}^\circ(g_{j,1} \otimes \cdots \otimes g_{j,n-1})_s \circ dW(g_{j,n})_s \\ &= \int_0^t J_{n-1}^\circ(g_{j,1} \otimes \cdots \otimes g_{j,n-1})_s dW(g_{j,n})_s + \frac{1}{2}\langle J_{n-1}^\circ(g_{j,1} \otimes \cdots \otimes g_{j,n-1}), W(g_{j,n})\rangle_t \\ &= \int_0^t J_{n-1}^\circ(g_{j,1} \otimes \cdots \otimes g_{j,n-1})_s g_{j,n}(s)^\top dX_s + \frac{1}{2}\int_0^t J_{n-2}^\circ(g_{j,1} \otimes \cdots \otimes g_{j,n-2})_s g_{j,n-1}(s)^\top d\langle X\rangle_s g_{j,n}(s) \end{aligned} \tag{7}$$

is also well-defined as stochastic integral (resp. Lebesgue-Stieltjes integral), since the integrand $s \mapsto J_{n-1}^\circ(g_{j,1} \otimes \cdots \otimes g_{j,n-1})_s g_n(s)$ is left-continuous and $\mathbb{F}$-adapted, thus $\mathbb{F}$-predictable and locally bounded.

Lemma 3.9. *Let X be a continuous semimartingale, let $n \in \mathbb{N}$, $p \in [1,\infty)$, and $t \in [0,T]$. Then, the following holds true:*

(i) *The operator $J_n^\circ(\cdot)_t : L^{np}(X)^{\otimes n} \to L^p(\Omega,\mathcal{F}_t,\mathbb{P})$ is linear.*

(ii) *$J_n^\circ(g)_t$ does (up to $\mathbb{P}$-null sets) not depend on the representation of $g \in L^{np}(X)^{\otimes n}$, i.e. if $g \in L^{np}(X)^{\otimes n}$ has representations $g^{(1)} = \sum_{j=1}^{m_1} g_{j,1}^{(1)} \otimes \cdots \otimes g_{j,n}^{(1)} \in L^{np}(X)^{\otimes n}$ and $g^{(2)} = \sum_{j=1}^{m_2} g_{j,1}^{(2)} \otimes \cdots \otimes g_{j,n}^{(2)} \in L^{np}(X)^{\otimes n}$, then $J_n^\circ\big(g^{(1)}\big)_t = J_n^\circ\big(g^{(2)}\big)_t$, $\mathbb{P}$-a.s.*

(iii) *There exists a constant $C_{n,p} > 0$ such that for every $g \in L^{np}(X)^{\otimes n}$, it holds that*

$$\mathbb{E}\left[\sup_{t\in[0,T]} |J_n^\circ(g)_t|^p\right]^{\frac{1}{p}} \leq C_{n,p}\|g\|_{L^{np}(X)^{\otimes n}},$$

with $C_{n,p} := \frac{3}{2}\prod_{k=1}^n \max\big(1, C_{\frac{np}{k}}\big) > 0$, where $C_r > 0$ denotes the constant of the classic upper Burkholder-Davis-Gundy inequality with exponent $r \in [1,\infty)$.

Moreover, we compute iterated Stratonovich integrals of symmetric tensors. To this end, we consider the vector subspace of symmetric tensors defined by

$$L^{np}_{\mathrm{sym}}(X)^{\otimes n} := \operatorname{span}\big\{g \in L^{np}(X)^{\otimes n} : \operatorname{sym}(g) = g\big\} \subseteq L^{np}(X)^{\otimes n}, \tag{8}$$

where $\operatorname{sym}(g) := \frac{1}{n!}\sum_{\sigma\in\mathcal{S}_n}\sum_{j=1}^m g_{j,\sigma(1)} \otimes \cdots \otimes g_{j,\sigma(n)}$ for $g = \sum_{j=1}^m g_{j,1} \otimes \cdots \otimes g_{j,n} \in L^{np}(X)^{\otimes n}$, with $\mathcal{S}_n$ denoting the set of permutations $\sigma : \{1,...,n\} \to \{1,...,n\}$.

Proposition 3.10. *Let X be a continuous semimartingale, let $n \in \mathbb{N}_0$, and $p \in [1,\infty)$. Then, for every $g = \sum_{j=1}^m g_{j,1} \otimes \cdots \otimes g_{j,n} \in L^{np}_{\mathrm{sym}}(X)^{\otimes n}$ and $t \in [0,T]$, it holds that $J_n^\circ(g)_t = \frac{1}{n!}\sum_{j=1}^m \prod_{k=1}^n W(g_{j,k})_t$, $\mathbb{P}$-a.s.*

4. CHAOS EXPANSION AND L^p-HEDGING

In this section, we show a chaos expansion of $L^p(\Omega,\mathcal{F}_T,\mathbb{P})$ into iterated Stratonovich integrals with respect to a given continuous semimartingale $X := \left(X_t^1,...,X_t^d\right)_{t\in[0,T]}$ having decomposition $X_t = X_0 + A_t + M_t$, $t\in[0,T]$, into a continuous $\mathbb{F}$-adapted process $A := \left(A_t^1,...,A_t^d\right)_{t\in[0,T]}$ of finite variation with $A_0 = 0$ and a continuous local martingale $M := \left(M_t^1,...,M_t^d\right)_{t\in[0,T]}$ with $M_0 = 0$.

As a consequence, every financial derivative $G\in L^p(\Omega,\mathcal{F}_T,\mathbb{P})$ can be approximated by finitely many iterated Stratonovich integrals. By combining this representation with a denseness result for $\mathbb{F}$-predictable processes, we further show that we can approximately solve the L^p-hedging problem.

4.1. Chaos expansion with iterated Stratonovich integrals.

In order to show that a chaos expansion of $L^p(\Omega,\mathcal{F}_T,\mathbb{P})$, we first prove that polynomials of X evaluated at different times are dense in $L^p(\Omega,\mathcal{F}_T,\mathbb{P})$. This is similar to [50, Proposition 2] for Lévy processes (see also [34]) and relies on the exponential integrability of X in Assumption 2.1. The proof can be found in Section 7.3.1.

Proposition 4.1. *Let X be a continuous semimartingale satisfying Assumption 2.1 and let $p\in[1,\infty)$. Then, the set of polynomials*

$$\operatorname{Pol}(X) := \operatorname{span}\left\{ \left(X_{t_1}^{i_1}\right)^{k_1}\cdots\left(X_{t_m}^{i_m}\right)^{k_m} : \begin{array}{c} m\in\mathbb{N},\ (i_1,...,i_m)\in\{1,...,d\}^m, \\ (k_1,...,k_m)\in\mathbb{N}_0^m,\ (t_1,...,t_m)\in[0,T]^m \end{array}\right\} \tag{9}$$

is dense in $L^p(\Omega,\mathcal{F}_T,\mathbb{P})$.

Now, we prove a chaos expansion of $L^p(\Omega,\mathcal{F}_T,\mathbb{P})$ into iterated Stratonovich integrals. More precisely, we show that the direct sum[4] of iterated Stratonovich integrals $J_n^\circ(g_n)_T$ with diagonal tensors

$$g_n\in L^{np}_{\text{diag}}(X)^{\otimes n} := \operatorname{span}\left\{g_0^{\otimes n} := g_0\otimes\cdots\otimes g_0 : g_0\in L^{np}(X)\right\}\subseteq L^{np}_{\text{sym}}(X)^{\otimes n}\subseteq L^{np}(X)^{\otimes n}$$

as integrands is dense in $L^p(\Omega,\mathcal{F}_T,\mathbb{P})$. The proof is given in Section 7.3.1.

Theorem 4.2 (Chaos expansion with iterated Stratonovich integrals)**.** *Let X be a continuous semimartingale satisfying Assumption 2.1 and let $p\in[1,\infty)$. Then, the following holds true:*

(i) *For $n, m_n\in\mathbb{N}$ and $g_n := \sum_{j=1}^{m_n} g_{n,j}^{\otimes n}\in L^{np}_{\text{diag}}(X)^{\otimes n}$ with $g_{n,j}\in L^{np}(X)$, $j=1,...,m_n$, we have*

$$J_n^\circ(g_n)_T = \int_0^T \left(\vartheta_t^{g_n}\right)^\top dX_t + \int_0^T \operatorname{tr}\left(\eta_t^{g_n} d\langle X\rangle_t\right), \qquad \mathbb{P}\text{-a.s.},$$

where the $\mathbb{R}^d$-valued process $\vartheta^{g_n} := \left(\vartheta_t^{g_n}\right)_{t\in[0,T]}$ and the $\mathbb{R}^{d\times d}$-valued process $\eta^{g_n} := \left(\eta_t^{g_n}\right)_{t\in[0,T]}$ are for every $t\in[0,T]$ given by

$$\vartheta_t^{g_n} := \sum_{j=1}^{m_n}\frac{W(g_{n,j})_t^{n-1}}{(n-1)!}g_{n,j}(t), \qquad \text{and} \qquad \eta_t^{g_n} := \frac{1}{2}\sum_{j=1}^{m_n}\frac{W(g_{n,j})_t^{n-1}}{(n-1)!}g_{n,j}(t)g_{n,j}(t)^\top.$$

(ii) *The direct sum $\bigoplus_{n\in\mathbb{N}_0}\left\{J_n^\circ(g_n)_T : g_n\in L^{np}_{\text{diag}}(X)^{\otimes n}\right\}$ is dense in $L^p(\Omega,\mathcal{F}_T,\mathbb{P})$, i.e. for every $G\in L^p(\Omega,\mathcal{F}_T,\mathbb{P})$ and $\varepsilon>0$ there exist $N\in\mathbb{N}$ and $g_n\in L^{np}_{\text{diag}}(X)^{\otimes n}$, $n=0,...,N$, such that*

$$\left\|G - \sum_{n=0}^N J_n^\circ(g_n)_T\right\|_{L^p(\mathbb{P})} < \varepsilon.$$

Remark 4.3. *Theorem 4.2 shows that iterated Stratonovich integrals are dense in $L^p(\Omega,\mathcal{F}_T,\mathbb{P})$ instead of an orthogonal decompositions of iterated Ito integrals in $L^2(\Omega,\mathcal{F}_T,\mathbb{P})$ proven in [16, 20, 32, 50, 63].*

4.2. L^p-hedging.

In this section, we consider the following L^p-hedging problem, where $p\in[1,\infty)$. For a given financial derivative $G\in L^p(\Omega,\mathcal{F}_T,\mathbb{P})$, we aim to find an optimal initial endowment $c\in\mathbb{R}$ and an optimal trading strategy $\theta := (\theta_t)_{t\in[0,T]}$ such that the L^p-hedging error $\left\|G - c - \int_0^T\theta_t^\top dX_t\right\|_{L^p(\mathbb{P})}$ is minimized. For $p=2$, this problem coincides with the quadratic hedging approaches in [52, 61], where sufficient conditions can be found for the existence of an optimal pair $c\in\mathbb{R}$ and $\theta := (\theta_t)_{t\in[0,T]}$.

[4] The direct sum of $(\mathcal{Z}_n)_{n\in\mathbb{N}_0}\subseteq L^p(\Omega,\mathcal{F}_T,\mathbb{P})$ is defined as $\bigoplus_{n=0}^\infty\mathcal{Z}_n := \left\{\sum_{n\in N} Z_n : N\subseteq\mathbb{N}_0 \text{ finite},\ Z_n\in\mathcal{Z}_n\right\}$.

We denote by $\Theta^p(X)$ the space of $\mathbb{F}$-predictable $\mathbb{R}^d$-valued processes $\theta := (\theta_t)_{t\in[0,T]}$ with

$$\|\theta\|_{\Theta^p(X)} := \sum_{i=1}^d \mathbb{E}\left[\left(\int_0^T \left|\theta_t^i dA_t^i\right|\right)^p\right]^{\frac{1}{p}} + \sum_{i=1}^d \mathbb{E}\left[\left(\int_0^T \left(\theta_t^i\right)^2 d\langle M^i\rangle_t\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} < \infty. \tag{10}$$

Note that the Kunita-Watanabe inequality and the Cauchy-Schwarz inequality ensure that for every $\theta \in \Theta^p(X)$ it holds that $\mathbb{E}\big[\big(\int_0^T |\theta_t^\top dA_t|\big)^p\big]^{1/p} + \mathbb{E}\big[\big(\int_0^T \theta_t^\top d\langle M\rangle_t \theta_t\big)^{p/2}\big]^{1/p} \leq \|\theta\|_{\Theta^p(X)}$.

Then, by combining the chaos expansion (Theorem 4.2) with a denseness result for $\mathbb{F}$-predictable processes, we are able to approximately solve the L^p-hedging problem. The proof is given in Section 7.3.2.

Theorem 4.4 (L^p-hedging)**.** *Let X be a continuous semimartingale satisfying Assumption 2.1 and let $p \in [1,\infty)$. Then, for every $G \in L^p(\Omega,\mathcal{F}_T,\mathbb{P})$ and $\varepsilon > 0$ there exist some $g_0 \in \mathbb{R}$ and $N, m_n \in \mathbb{N}$ as well as $g_{n,j,0}, g_{n,j,1} \in L^{np}(X)$, $n = 1,\dots,N$ and $j = 1,\dots,m_n$, such that*

$$\left\|G - g_0 - \int_0^T \left(\vartheta_t^{g_{1:N}}\right)^\top dX_t\right\|_{L^p(\mathbb{P})} \leq \varepsilon + \inf_{(c,\theta)\in\mathbb{R}\times\Theta^p(X)} \left\|G - c - \int_0^T \theta_t^\top dX_t\right\|_{L^p(\mathbb{P})},$$

where the $\mathbb{R}^d$-valued process $\vartheta^{g_{1:N}} := \left(\vartheta_t^{g_{1:N}}\right)_{t\in[0,T]}$ is for every $t \in [0,T]$ given by

$$\vartheta_t^{g_{1:N}} := \sum_{n=1}^N \sum_{j=1}^{m_n} \frac{W(g_{n,j,0})_t^{n-1}}{(n-1)!} g_{n,j,1}(t).$$

In the following, we use (possibly random) neural networks to learn the integrands $g_n \in L^{np}(X)^{\otimes n}$, $n = 0,\dots,N$, in the chaos expansion (Theorem 4.2) and the functions $g_{n,j,0}, g_{n,j,1} \in L^{np}(X)$, $n = 1,\dots,N$ and $j = 1,\dots,m_n$ in the approximation of the L^p-hedging problem (Theorem 4.4).

5. Universal approximation of financial derivatives

In this section, we approximate any given financial derivative $G \in L^p(\Omega,\mathcal{F}_T,\mathbb{P})$ in two steps: first by using the chaos expansion (Theorem 4.2) and then by replacing the deterministic integrands by (possibly random) neural networks (Theorem 5.4+5.9). We also refer to Algorithm 1 for the procedure.

5.1. Fully trained neural networks.

Inspired by the functionality of a human brain, neural networks were introduced in [45]. Here, we generalize them to our tensor-valued setting.

Definition 5.1. *For $p \in [1,\infty)$ and $n, m \in \mathbb{N}$, a diagonal tensor $\varphi = \sum_{j=1}^m y_j \varphi_j^{\otimes n} \in L^{np}_{\text{diag}}(X)^{\otimes n}$ is called a* (tensor-valued) neural network *if for every $j = 1,\dots,m$ it holds that*

$$[0,T] \ni t \quad \mapsto \quad \varphi_j(t) = \rho(a_j t + b_j) \in \mathbb{R}^d, \tag{11}$$

for some activation function *$\rho \in C(\mathbb{R})$ that is applied componentwise in (11). Hereby, $a_1,\dots,a_m \in \mathbb{R}^d$ are* weights*, $b_1,\dots,b_m \in \mathbb{R}^d$ are* biases*, and $y_1,\dots,y_m \in \mathbb{R}$ are* linear readouts*. Moreover, we define $\mathcal{NN}^\rho_{d,n}$ as the set of all tensor-valued neural networks, whereas for $n = 0$ we set $\mathcal{NN}^\rho_{d,0} := \mathbb{R}$.*

Remark 5.2. *Since $\rho \in C(\mathbb{R})$ is continuous, Lemma 3.2 (ii) ensures that $\varphi_j \in L^p(X)$ for all $j = 1,\dots,m$. Therefore, every tensor-valued neural network $\varphi \in \mathcal{NN}^\rho_{d,n}$ is a diagonal tensor in $L^{np}_{\text{diag}}(X)^{\otimes n}$.*

Now, we use the universal approximation property of neural networks (see, e.g., [12, 29, 53]) to approximate any diagonal tensor in $L^{np}_{\text{diag}}(X)^{\otimes n}$ by a tensor-valued neural network. To this end, we assume that the activation function $\rho \in C(\mathbb{R})$ is non-polynomial, i.e. $\rho \in C(\mathbb{R})$ is almost everywhere algebraically not equal to a polynomial (see also [41]). The proof can be found in Section 7.4.1.

Proposition 5.3. *Let X be a continuous semimartingale satisfying Assumption 2.1, let $n \in \mathbb{N}_0$, $p \in [1,\infty)$, and let $\rho \in C(\mathbb{R})$ be non-polynomial. Then, $\mathcal{NN}^\rho_{d,n}$ is dense in $L^{np}_{\text{diag}}(X)^{\otimes n}$, i.e. for every $g \in L^{np}_{\text{diag}}(X)^{\otimes n}$ and $\varepsilon > 0$ there exists some $\varphi \in \mathcal{NN}^\rho_{d,n}$ such that $\|g - \varphi\|_{L^{np}(X)^{\otimes n}} < \varepsilon$.*

Hence, by using the chaos expansion (Theorem 4.2) together with Proposition 5.3, we obtain the following universal approximation theorem for financial derivatives. The proof is given in Section 7.4.1.

Theorem 5.4 (Universal approximation)**.** *Let X be a continuous semimartingale satisfying Assumption 2.1, let $p \in [1,\infty)$, and let $\rho \in C(\mathbb{R})$ be non-polynomial. Then, $\bigoplus_{n\in\mathbb{N}_0} \left\{J_n^\circ(\varphi_n)_T : \varphi_n \in \mathcal{NN}^\rho_{d,n}\right\}$ is dense in $L^p(\Omega,\mathcal{F}_T,\mathbb{P})$, i.e. for every $G \in L^p(\Omega,\mathcal{F}_T,\mathbb{P})$ and $\varepsilon > 0$ there exists some $N \in \mathbb{N}$ and $\varphi_n \in \mathcal{NN}^\rho_{d,n}$, $n = 0,\dots,N$, such that $\left\|G - \sum_{n=0}^N J_n^\circ(\varphi_n)_T\right\|_{L^p(\mathbb{P})} < \varepsilon$.*

5.2. **Random neural networks.** Inspired by the works on extreme learning machines and random feature learning (see [23, 24, 26, 30, 46, 55]), we now consider random neural networks, instead of fully trained neural networks, where the weights and biases are now randomly initialized. In this case, only the linear readout needs to be trained, which can be performed efficiently, e.g., by the least squares method.

To this end, we impose the following condition on the activation function $\rho \in C(\mathbb{R})$ and the random initialization of the weights and biases $\widetilde{a}_1, \widetilde{b}_1 : \widetilde{\Omega} \to \mathbb{R}^d$ defined on another probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$.

Assumption 5.5. *Let $\widetilde{a}_1, \widetilde{b}_1 : \widetilde{\Omega} \to \mathbb{R}^d$ be two random vectors defined on a (possibly different) probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ such that for every $(a,b) \in \mathbb{R}^d \times \mathbb{R}^d$ and $r > 0$ it holds that $\widetilde{\mathbb{P}}[\{\widetilde{\omega} \in \widetilde{\Omega} : \|(\widetilde{a}_1(\widetilde{\omega}), \widetilde{b}_1(\widetilde{\omega})) - (a,b)\| < r\}] > 0$. Moreover, let $\rho \in C(\mathbb{R})$ such that for every $q \in [1,\infty)$ we have*

$$\widetilde{\mathbb{E}}\left[\left\|\rho\left(\widetilde{a}_1 \cdot + \widetilde{b}_1\right)\right\|_\infty^q\right] := \widetilde{\mathbb{E}}\left[\sup_{t\in[0,T]}\left\|\rho\left(\widetilde{a}_1 t + \widetilde{b}_1\right)\right\|^q\right] < \infty.$$

Example 5.6. *Assumption 5.5 is satisfied for example if $\rho \in C(\mathbb{R})$ is bounded (e.g., sigmoid function $\rho(s) = \frac{1}{1+\exp(-s)}$) or if there exist $C, c > 0$ such that $|\rho(s)| \leq C(1+|s|^c)$, for all $s \in \mathbb{R}$, and $\widetilde{a}_1$ and $\widetilde{b}_1$ have moments of all orders (e.g., ReLU function $\rho(s) = \max(s,0)$ and normally distributed $\widetilde{a}_1, \widetilde{b}_1$).*

Moreover, we assume that $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ supports two sequences of independent and identically distributed (i.i.d.) random variables $(\widetilde{a}_j)_{j\in\mathbb{N}} \sim \widetilde{a}_1$ and $(\widetilde{b}_j)_{j\in\mathbb{N}} \sim \widetilde{b}_1$ used for the random weights and biases, and define the σ-algebra $\widetilde{\mathcal{F}}_{\mathrm{rand}} := \sigma(\{\widetilde{a}_j, \widetilde{b}_j : j \in \mathbb{N}\})$. Now, we can introduce random neural networks.

Definition 5.7. *For $p \in [1,\infty)$ and $n, m \in \mathbb{N}$, we call $\widetilde{\Omega} \ni \widetilde{\omega} \mapsto \widetilde{\varphi}(\widetilde{\omega}) := \sum_{j=1}^m \widetilde{y}_j(\widetilde{\omega}) \varphi_j(\widetilde{\omega})^{\otimes n} \in L^{np}_{\mathrm{diag}}(X)^{\otimes n}$ a* (tensor-valued) random neural network *if for every $\widetilde{\omega} \in \widetilde{\Omega}$ and $j = 1, ..., m$ it holds that*

$$[0,T] \ni t \quad \mapsto \quad \widetilde{\varphi}_j(\widetilde{\omega})(t) = \rho\left(\widetilde{a}_j(\widetilde{\omega}) t + \widetilde{b}_j(\widetilde{\omega})\right) \in \mathbb{R}^d, \tag{12}$$

for some activation function $\rho \in C(\mathbb{R})$ *that is applied componentwise in (12). Hereby, $(\widetilde{a}_j)_{j\in\mathbb{N}} \overset{i.i.d.}{\sim} \widetilde{a}_1$ are the* random weights, *$(\widetilde{b}_j)_{j\in\mathbb{N}} \overset{i.i.d.}{\sim} \widetilde{b}_1$ are the* random biases, *and $\widetilde{y}_1, ..., \widetilde{y}_m : \widetilde{\Omega} \to \mathbb{R}$ are the* linear readouts *that are assumed to be $\widetilde{\mathcal{F}}_{\mathrm{rand}}/\mathcal{B}(\mathbb{R})$-measurable and $\widetilde{\mathbb{P}}$-a.s. bounded. Moreover, we denote by $\mathcal{RN}^\rho_{d,n}$ the set of tensor-valued random neural networks, whereas for $n = 0$ we set $\mathcal{RN}^\rho_{d,0} := \mathbb{R}$.*

In Lemma 7.6, we show that every random neural network $\widetilde{\varphi} \in \mathcal{RN}^\rho_{d,n}$ is well-defined in the Bochner space $L^r(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}}; \overline{L^{np}(X)^{\otimes n}})$ (see, e.g., [31] for more details), where $\overline{L^{np}(X)^{\otimes n}}$ denotes the completion of $L^{np}(X)^{\otimes n}$. Now, we apply the strong law of large numbers for Banach space-valued random variables to obtain the following universal approximation result for random neural networks.

Proposition 5.8. *Let X be a continuous semimartingale satisfying Assumption 2.1, let $n \in \mathbb{N}_0$, $p, r \in [1,\infty)$, $g \in L^{np}(X)^{\otimes n}$, let $\rho \in C(\mathbb{R})$ be non-polynomial, and let Assumption 5.5 hold. Then, for every $\varepsilon > 0$ there exists some $\widetilde{\varphi} \in \mathcal{RN}^\rho_{d,n}$ such that $\widetilde{\mathbb{E}}\left[\|g - \widetilde{\varphi}(\cdot)\|^r_{L^{np}(X)^{\otimes n}}\right]^{1/r} < \varepsilon$.*

Proposition 5.8 implies the following approximation result for financial derivatives, which is similar to Theorem 5.4 but now with random neural networks. The proof can be found in Section 7.4.2.

Theorem 5.9 (Random universal approximation)**.** *Let X be a continuous semimartingale satisfying Assumption 2.1, let $p, r \in [1,\infty)$, let $\rho \in C(\mathbb{R})$ be non-polynomial, and let Assumption 5.5 hold. Then, for every $G \in L^p(\Omega, \mathcal{F}_T, \mathbb{P})$ and $\varepsilon > 0$ there exists $N \in \mathbb{N}$ and $\widetilde{\varphi}_n \in \mathcal{RN}^\rho_{d,n}$, $n = 0, ..., N$, such that $\left(\widetilde{\omega} \mapsto \sum_{n=0}^N J_n^\circ(\widetilde{\varphi}_n(\widetilde{\omega}))_T\right) \in L^r(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}}; L^p(\Omega, \mathcal{F}_T, \mathbb{P}))$ satisfies $\widetilde{\mathbb{E}}\left[\left\|G - \sum_{n=0}^N J_n^\circ(\widetilde{\varphi}_n(\widetilde{\omega}))_T\right\|^r_{L^p(\mathbb{P})}\right]^{1/r} < \varepsilon$.*

Remark 5.10. *The universal approximation results in Theorem 5.4 (for fully trained neural networks) and in Theorem 5.9 (for random neural networks) extend the following results in the literature:*

(i) To learn financial derivatives, the signature *has been applied in [5, 10, 11, 43, 44]. The signature stems from rough path theory (see [42]) and satisfies a similar universality as neural networks but for path space functionals. Since the signature of a semimartingale is given as iterated (Ito) integrals with constant integrand equal to one, Theorem 5.4 can be related to these results.*

(ii) The random universal approximation result in Proposition 5.8 is similar to [23, 24, 46].

5.3. **L^p-hedging.** For $p \in [1,\infty)$, we now consider again the L^p-hedging problem from Section 4.2. However, instead of arbitrary $L^{np}(X)$-functions in Theorem 4.4, we now use (possibly random) neural networks, which are still able to approximately solve the L^p-hedging problem. First, we provide the result for fully trained neural networks, whose proof is given in Section 7.4.3.

Theorem 5.11 (L^p-hedging with fully trained neural networks)**.** *Let X be a continuous semimartingale satisfying Assumption 2.1, let $p \in [1,\infty)$, and let $\rho \in C(\mathbb{R})$ be non-polynomial. Then, for every $G \in L^p(\Omega,\mathcal{F}_T,\mathbb{P})$ and $\varepsilon > 0$ there exist some $\varphi_0 \in \mathbb{R}$ and $N, m_n \in \mathbb{N}$ as well as $\varphi_{n,j,0}, \varphi_{n,j,1} \in \mathcal{NN}^\rho_{d,1}$, $n = 1,...,N$ and $j = 1,...,m_n$, such that*

$$\left\| G - \varphi_0 - \int_0^T \left(\vartheta_t^{\varphi_{1:N}}\right)^\top dX_t \right\|_{L^p(\mathbb{P})} \leq \varepsilon + \inf_{(c,\theta)\in\mathbb{R}\times\Theta^p(X)} \left\| G - c - \int_0^T \theta_t^\top dX_t \right\|_{L^p(\mathbb{P})},$$

where $\vartheta_t^{\varphi_{1:N}} := \sum_{n=1}^N \sum_{j=1}^{m_n} \frac{W(\varphi_{n,j,0})_t^{n-1}}{(n-1)!} \varphi_{n,j,1}(t)$, for $t \in [0,T]$.

Next, we use random neural networks to approximately solve the L^p-hedging problem.

Theorem 5.12 (L^p-hedging with random neural networks)**.** *Let X be a continuous semimartingale satisfying Assumption 2.1, let $p, r \in [1,\infty)$, let $\rho \in C(\mathbb{R})$ be non-polynomial, and let Assumption 5.5 hold. Then, for every $G \in L^p(\Omega,\mathcal{F}_T,\mathbb{P})$ and $\varepsilon > 0$ there exist some $\widetilde{\varphi}_0 \in \mathbb{R}$ and $N, m_n \in \mathbb{N}$ as well as $\widetilde{\varphi}_{n,j,0}, \widetilde{\varphi}_{n,j,1} \in \mathcal{RN}^\rho_{d,1}$, $n = 1,...,N$ and $j = 1,...,m_n$, such that*

$$\widetilde{\mathbb{E}}\left[\left\| G - \widetilde{\varphi}_0 - \int_0^T \left(\vartheta_t^{\widetilde{\varphi}_{1:N}(\cdot)}\right)^\top dX_t \right\|_{L^p(\mathbb{P})}^r \right]^{\frac{1}{r}} \leq \varepsilon + \inf_{(c,\theta)\in\mathbb{R}\times\Theta^p(X)} \left\| G - c - \int_0^T \theta_t^\top dX_t \right\|_{L^p(\mathbb{P})},$$

where $\vartheta_t^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})} := \sum_{n=1}^N \sum_{j=1}^{m_n} \frac{W(\widetilde{\varphi}_{n,j,0}(\widetilde{\omega}))_t^{n-1}}{(n-1)!} \widetilde{\varphi}_{n,j,1}(\widetilde{\omega})(t)$, for $t \in [0,T]$.

Hence, by using (possibly random) neural networks, we are indeed able to approximately solve the L^p-hedging problem, which coincides for $p = 2$ with the quadratic hedging problem described in [52, 61].

6. Numerical examples

In this section, we numerically solve the L^p-hedging problem described in Section 4.2+5.3. To this end, we assume that X is a continuous semimartingale satisfying Assumption 2.1. Then, for a given financial derivative $G \in L^p(\Omega,\mathcal{F}_T,\mathbb{P})$, we follow Algorithm 1 to learn the trading strategy from Theorem 5.12.

In the following numerical experiments[5], we use random neural networks to learn the approximating L^p-optimal strategy of a financial derivative within a few seconds. To this end, we generate $L = 10^5$ sample paths of the semimartingale X, discretized over an equidistant time grid $(t_k)_{k=0,...,K}$ of $K = 500$ time points, which are split up into 80%/20% for training and testing. Then, for each order of the chaos expansion $n = 0,...,N$, we initialize random neurons of size $m_n = 50$ with non-polynomial activation function $\rho(s) = \tanh(s)$.

In order to compare our results with the quadratic hedging approach in [52, 61], we choose $p = 2$. Then, for a given financial derivative $G \in L^2(\Omega,\mathcal{F}_T,\mathbb{P})$ and different $N = 0,...,6$, we learn the approximating L^p-optimal strategy $\vartheta^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})}$ from Theorem 5.12. The following numerical examples empirically demonstrate that the pseudo-optimal L^2-strategy of $G \in L^2(\Omega,\mathcal{F}_T,\mathbb{P})$, as defined in [61, p. 17], can indeed be approximated by $\vartheta^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})}$.

6.1. **European call option with Brownian motion.** In the first example, we consider a one-dimensional Brownian motion $X := (X_t)_{t\in[0,T]}$, which is a continuous semimartingale satisfying Assumption 2.1. Then, we approximately solve the L^2-hedging problem for the European call option

$$G = \max(X_T - K_{\text{str}}, 0), \tag{14}$$

where $K_{\text{str}} \in \mathbb{R}$ denotes the strike price. Since the market is complete and $\mathbb{P}$ is a (local) martingale measure, the L^2-optimal strategy is the replication strategy under $\mathbb{P}$. Hence, by using the Clark-Ocone formula in [7, 51] (see also [15, Theorem 4.1]), the true hedging strategy $\theta = (\theta_t)_{t\in[0,T]}$ of G is for every $t \in [0,T]$ given by $\theta_t = \mathbb{E}[D_t G|\mathcal{F}_t]$, where $D_t G$ denotes the Malliavin derivative of G

[5] All numerical experiments have been implemented in `Python` on an average laptop (Lenovo ThinkPad X13 Gen2a with Processor AMD Ryzen 7 PRO 5850U and Radeon Graphics, 1901 Mhz, 8 Cores, 16 Logical Processors). The code can be found under the following link: https://github.com/psc25/ChaoticHedging

Algorithm 1: Learning the L^p-optimal strategy of a financial derivative.

Input: Financial derivative $G \in L^p(\Omega, \mathcal{F}_T, \mathbb{P})$, for some $p \in [1, \infty)$, and $K, L, N, m_1, ..., m_N \in \mathbb{N}$.
Output: Approximating L^p-optimal strategy $\vartheta^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})}$ of $G \in L^p(\Omega, \mathcal{F}_T, \mathbb{P})$ from Theorem 5.12.

1 Generate L discretized sample paths $(X_{t_k}(\omega_l))_{k=0,...,K}$ of X, $l = 1, ..., L$, where $0 = t_0 < ... < t_K = T$ is a partition of $[0,T]$, and compute for every $l = 1, ..., L$ the corresponding payoff $G(\omega_l)$.

2 Let $\rho \in C(\mathbb{R})$ be non-polynomial and initialize the weights $(\widetilde{a}_{n,j,0})_{n,j\in\mathbb{N}}, (\widetilde{a}_{n,j,1})_{n,j\in\mathbb{N}} \overset{\text{i.i.d.}}{\sim} \widetilde{a}_1$ and biases $(\widetilde{b}_{n,j,0})_{n,j\in\mathbb{N}}, (\widetilde{b}_{n,j,1})_{n,j\in\mathbb{N}} \overset{\text{i.i.d.}}{\sim} \widetilde{b}_1$ randomly, i.e. fix some $\widetilde{\omega} \in \widetilde{\Omega}$. Hereby, Assumption 5.5 should hold.

3 Let $\widetilde{\varphi}_0 \in \mathbb{R}$ and define for every $n = 1, ..., N$ and $j = 1, ..., m_n$ the random neurons

$$
\begin{aligned}
[0,T] \ni t \quad &\mapsto \quad \widetilde{\varphi}_{n,j,0}(\widetilde{\omega})(t) := \rho\left(\widetilde{a}_{n,j,0}(\widetilde{\omega})t + \widetilde{b}_{n,j,0}(\widetilde{\omega})\right) \in \mathbb{R}^d, \\
[0,T] \ni t \quad &\mapsto \quad \widetilde{\varphi}_{n,j,1}(\widetilde{\omega})(t) := \rho\left(\widetilde{a}_{n,j,1}(\widetilde{\omega})t + \widetilde{b}_{n,j,1}(\widetilde{\omega})\right) \in \mathbb{R}^d.
\end{aligned}
$$

4 Compute for every $n = 1, ..., N$, $j = 1, ..., m_n$, $k = 0, ..., K$, and $l = 1, ..., L$ the stochastic integrals $W(\widetilde{\varphi}_{n,j,0}(\widetilde{\omega}))_{t_k}(\omega_l)$ and $W(\widetilde{\varphi}_{n,j,1}(\widetilde{\omega}))_{t_k}(\omega_l)$, e.g., by using Ito's formula, i.e.

$$
\begin{aligned}
W(\widetilde{\varphi}_{n,j,0}(\widetilde{\omega}))_{t_k} &:= \int_0^{t_k} \widetilde{\varphi}_{n,j,0}(\widetilde{\omega})(s)^\top dX_s \\
&= \widetilde{\varphi}_{n,j,0}(\widetilde{\omega})(t_k)^\top X_{t_k} - \widetilde{\varphi}_{n,j,0}(\widetilde{\omega})(0)^\top X_0 - \int_0^{t_k} \widetilde{\varphi}_{n,j,0}(\widetilde{\omega})'(s)^\top X_s \, ds
\end{aligned}
$$

and similarly for $W(\widetilde{\varphi}_{n,j,1}(\widetilde{\omega}))_{t_k} := \int_0^{t_k} \widetilde{\varphi}_{n,j,1}(\widetilde{\omega})(s)^\top dX_s$, where $\rho \in C(\mathbb{R})$ must now be differentiable.

5 Let $y_{n,j} \in \mathbb{R}$, $n = 1, ..., N$ and $j = 1, ..., m_n$, and initialize the trading strategy $\vartheta^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})} := \left(\vartheta_t^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})}\right)_{t\in[0,T]}$ for every $t \in [0,T]$ by

$$
\vartheta_t^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})} := \sum_{n=1}^N \sum_{j=1}^{m_n} y_{n,j} \frac{W(\widetilde{\varphi}_{n,j,0}(\widetilde{\omega}))_t^{n-1}}{(n-1)!} \widetilde{\varphi}_{n,j,1}(\widetilde{\omega})(t).
$$

6 Minimize the empirical L^p-hedging error over $\widetilde{\varphi}_0 \in \mathbb{R}$ and $y_{n,j} \in \mathbb{R}$, $n = 1, ..., N$ and $j = 1, ..., m_n$, i.e.

$$
\inf_{\widetilde{\varphi}_0, y_{n,j} \in \mathbb{R}} \left(\sum_{l=1}^L \left| G(\omega_l) - \widetilde{\varphi}_0 - \left(\int_0^T \left(\vartheta_t^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})} \right)^\top dX_t \right)(\omega_l) \right|^p \right)^{\frac{1}{p}}, \tag{13}
$$

where the stochastic integral

$$
\int_0^T \left(\vartheta_t^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})} \right)^\top dX_t = \sum_{n=1}^N \sum_{j=1}^{m_n} y_{n,j} \int_0^T \frac{W(\widetilde{\varphi}_{n,j,0}(\widetilde{\omega}))_t^{n-1}}{(n-1)!} \widetilde{\varphi}_{n,j,1}(\widetilde{\omega})(t)^\top dX_t
$$

is approximated, e.g., by an Euler-Maruyama scheme.

7 Update the parameters $\widetilde{\varphi}_0 \in \mathbb{R}$ and $y_{n,j} \in \mathbb{R}$, $n = 1, ..., N$ and $j = 1, ..., m_n$, to the ones minimizing (13) and return the approximating L^p-optimal strategy $\vartheta^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})} := \left(\vartheta_t^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})}\right)_{t\in[0,T]}$.

Fully trained neural networks: *If fully trained neural networks are used instead of random neural networks, we omit the $\widetilde{\omega}$-dependence in Algorithm 1 and replace Step 3+6 by the following steps:*

3 Let $\varphi_0 \in \mathbb{R}$ and define for every $n = 1, ..., N$ and $j = 1, ..., m_n$ the fully trained neurons

$$
\begin{aligned}
[0,T] \ni t \quad &\mapsto \quad \varphi_{n,j,0}(t) := \rho\left(a_{n,j,0}t + b_{n,j,0}\right) \in \mathbb{R}^d, \\
[0,T] \ni t \quad &\mapsto \quad \varphi_{n,j,1}(t) := \rho\left(a_{n,j,1}t + b_{n,j,1}\right) \in \mathbb{R}^d,
\end{aligned}
$$

for some $a_{n,j,0}, a_{n,j,1}, b_{n,j,0}, b_{n,j,1} \in \mathbb{R}^d$, $n = 1, ..., N$ and $j = 1, ..., m_n$.

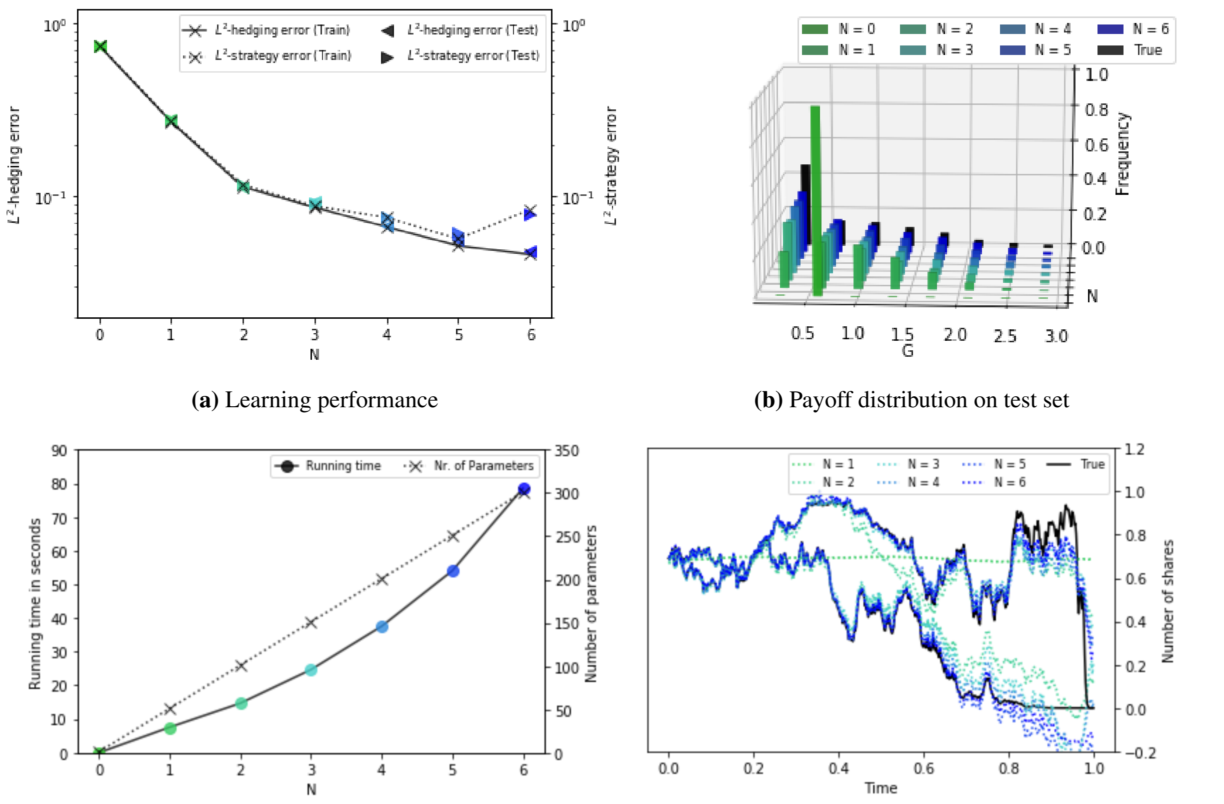
6 Minimize (13) over $\varphi_0 \in \mathbb{R}$ as well as $a_{n,j,0}, a_{n,j,1}, b_{n,j,0}, b_{n,j,1}, y_{n,j} \in \mathbb{R}$, $j = 1, ..., m_n$ and $n = 1, ..., N$. Then, $\vartheta^{\varphi_{1:N}} := \left(\vartheta_t^{\varphi_{1:N}}\right)_{t\in[0,T]}$ is the approximating L^p-optimal strategy from Theorem 5.11.

Comparison: *Random neural networks outperform fully trained neural networks in terms of computational efficiency and stability. Indeed, fully trained neural networks require an iterative algorithm for training (e.g., stochastic gradient descent, see [25]), whereas only the linear readout of a random neural network needs to be trained (e.g., least squares method). Moreover, the corresponding optimization problem for training a random neural network is convex, which is not the case for fully trained neural networks (see [46]).*

(see [49, Definition 1.2.1]). Using the chain rule of the Malliavin derivative in [49, Proposition 1.2.4], it follows for every $t \in [0,T)$ that $D_t G = \mathbb{1}_{[K_{\text{str}},\infty)}(X_T)$ and thus

$$\theta_t = \mathbb{E}\left[\mathbb{1}_{[K_{\text{str}},\infty)}(X_T)\big|\mathcal{F}_t\right] = \int_{K_{\text{str}}}^{\infty} \frac{1}{\sqrt{2\pi(T-t)}} e^{-\frac{(x-X_t)^2}{2(T-t)}} dx = \Phi\left(\frac{X_t - K_{\text{str}}}{\sqrt{T-t}}\right), \tag{15}$$

where $\Phi : \mathbb{R} \to [0,1]$ denotes the cumulative distribution function of the standard normal distribution. For the numerical experiment in Figure 1, we choose the parameters $T = 1$ and $K_{\text{str}} = -0.5$.



(a) Learning performance

(b) Payoff distribution on test set

(c) Running time and number of parameters

(d) θ and $\vartheta^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})}$ for two samples of test set

Figure 1. Learning the L^2-optimal strategy of the European call option G defined in (14), for $N = 0, ..., 6$. In (a), the $L^2(\mathbb{P})$-hedging error (13) and the empirical $L^2(dt \otimes \mathbb{P})$-strategy error $\left(\frac{T}{KL}\sum_{l=1}^{L}\sum_{k=0}^{K}\left|\theta_{t_k}(\omega_l) - \vartheta_{t_k}^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})}(\omega_l)\right|^2\right)^{1/2}$ are displayed, where θ is defined in (15). In (b), the distributions of G (label "True") and $\widetilde{\varphi}_0 + \int_0^T \vartheta_t^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})} dX_t$ (label "$N = 0, ..., 6$") are depicted on the test set. In (c), the running time[4] and the number of estimated parameters are shown. In (d), we compare θ (label "True") and $\vartheta^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})}$ (label "$N = 1, ..., 6$") on two samples of the test set.

6.2. **Asian put option in a Vašiček model with stochastic volatility.** In the second example, we consider a Vašiček model with stochastic volatility, i.e. the price process $X := (X_t)_{t\in[0,T]}$ follows the SDE

$$\begin{cases} dX_t = \kappa\,(\mu - X_t)\,dt + \sqrt{\sigma_t} dB_t, \\ d\sigma_t \;= \alpha(\beta - \sigma_t)dt + \xi\sqrt{\sigma_t} d\widetilde{B}_t, \end{cases} \qquad t \in [0,T] \tag{16}$$

with initial values $(X_0, \sigma_0) \in \mathbb{R} \times (0,\infty)$ and parameters $\kappa, \mu, \alpha, \beta, \xi > 0$. Hereby, the Brownian motions $B := (B_t)_{t\in[0,T]}$ and $\widetilde{B} := \left(\widetilde{B}_t\right)_{t\in[0,T]}$ are correlated with parameter $\rho \in [-1,1]$, i.e. $\widetilde{B}_t = \rho B_t + \sqrt{1-\rho^2} B_t^\perp$ for all $t \in [0,T]$ and another Brownian motion $B^\perp := \left(B_t^\perp\right)_{t\in[0,T]}$ which is independent of B. Since $\widetilde{X} := (X_t, \sigma_t)_{t\in[0,T]}^\top$ is a polynomial diffusion with diffusion coefficient

[4] The running time corresponds to the total execution time of Algorithm 1 including the computation of all inputs, solving the least squares problem (taking less than one second), and the calculation of the approximating L^2-optimal strategy.

$a:\mathbb{R}\times(0,\infty)\to\mathbb{S}^2_+$ of linear growth, Example 2.3 ensures that $\widetilde{X}$ satisfies Assumption 2.1, which thus also applies to X. In this setting, we approximately solve the L^2-hedging problem for the Asian put option

$$G=\max\left(K_{\text{str}}-\frac{1}{T}\int_0^T X_t dt, 0\right), \tag{17}$$

where $K_{\text{str}}\in\mathbb{R}$ denotes the strike price.

In order to compute the L^2-optimal strategy in the sense of quadratic hedging (see [52, 61]), we first observe that the market price of risk $t\mapsto\lambda_t:=\sigma_t^{-1}\kappa(\mu-X_t)$ induces the minimal equivalent local martingale measure $\mathbb{Q}\sim\mathbb{P}$ whose density $\frac{d\mathbb{Q}}{d\mathbb{P}}:=Z_T$ is the terminal value of the SDE $dZ_t=-Z_t\lambda_t\sqrt{\sigma_t}dB_t$, with $Z_0=1$. Hence, X is by the Girsanov theorem a local martingale under $\mathbb{Q}$ satisfying the SDE $dX_t=\sqrt{\sigma_t}dB_t^{\mathbb{Q}}$, where $B^{\mathbb{Q}}:=\left(B_t^{\mathbb{Q}}\right)_{t\in[0,T]}$ is a Brownian motion under $\mathbb{Q}$. Thus, [61, Theorem 3.5] shows that $t\mapsto V_t^{\mathbb{Q}}:=\mathbb{E}^{\mathbb{Q}}[G|\mathcal{F}_t]$ is the value process of the pseudo-optimal L^2-strategy in the sense of [61, p. 17]. For the calculation of $\mathbb{E}^{\mathbb{Q}}[G|\mathcal{F}_t]$, we use the notation $I_{r,s}:=\frac{1}{T}\int_r^s X_t dt$, the Fourier arguments in [4] together with the put-call parity, and the affine transform formulas in [36, Corollary 4.11] and [17, Section 2.3] to conclude for every $t\in[0,T]$ that

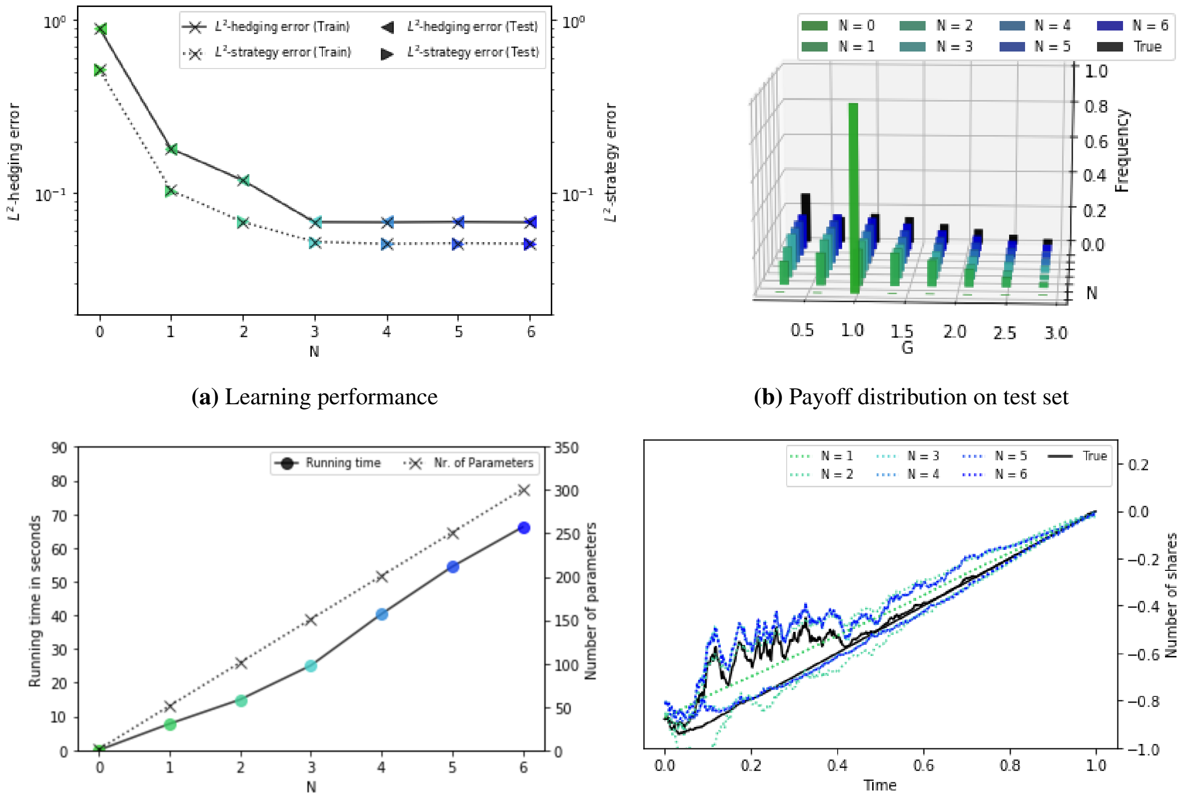
$$\begin{aligned}
V_t^{\mathbb{Q}} &= \mathbb{E}^{\mathbb{Q}}[G|\mathcal{F}_t]=\mathbb{E}^{\mathbb{Q}}\left[\max\left(K_{\text{str}}-I_{0,t}-I_{t,T},0\right)|\mathcal{F}_t\right] \\
&= (K_{\text{str}}-I_{0,t})\mathbb{E}^{\mathbb{Q}}\left[\mathbb{1}_{(-\infty,K_{\text{str}}-I_{0,t}]}(I_{t,T})\middle|\mathcal{F}_t\right]-\mathbb{E}^{\mathbb{Q}}\left[I_{t,T}\mathbb{1}_{(-\infty,K_{\text{str}}-I_{0,t}]}(I_{t,T})\middle|\mathcal{F}_t\right] \\
&= (K_{\text{str}}-I_{0,t})\left(\frac{1}{2}-\int_0^\infty \operatorname{Re}\left(\frac{\exp\left(-\mathbf{i}u(K_{\text{str}}-I_{0,t})\right)\mathbb{E}\left[\exp\left(\mathbf{i}uI_{t,T}\right)\middle|\mathcal{F}_t\right]}{\mathbf{i}\pi u}\right)du\right) \\
&\quad-\left(\frac{\mathbb{E}^{\mathbb{Q}}[I_{t,T}]}{2}+\frac{1}{2}\int_0^\infty \operatorname{Re}\left(\frac{\exp\left(-\mathbf{i}u(K-I_{0,t})\right)\mathbb{E}\left[\mathbf{i}I_{t,T}\exp\left(\mathbf{i}uI_{t,T}\right)\middle|\mathcal{F}_t\right]}{\mathbf{i}\pi u}\right)du\right) \\
&= (K_{\text{str}}-I_{0,t})\left(\frac{1}{2}-\int_0^\infty \operatorname{Re}\left(\frac{\exp\left(-\mathbf{i}u(K_{\text{str}}-I_{0,t})\right)\exp\left(\phi_{T-t}^{(0)}(u)^\top\widetilde{X}_t\right)}{\mathbf{i}\pi u}\right)du\right) \\
&\quad-\frac{T-t}{2T}X_t-\int_0^\infty \operatorname{Re}\left(\frac{\exp\left(-\mathbf{i}u(K_{\text{str}}-I_{0,t})\right)\exp\left(\phi_{T-t}^{(0)}(u)^\top\widetilde{X}_t\right)\phi_{T-t}^{(1)}(u)^\top\widetilde{X}_t}{\pi u}\right)du,
\end{aligned}$$

where $\phi_h^{(0)},\phi_h^{(1)}:\mathbb{C}\to\mathbb{C}^2$ satisfy the Ricatti equations $\frac{\partial}{\partial h}\phi_h^{(0)}(u)=\left(\mathbf{i}u,\frac{1}{2}\phi_h^{(0)}(u)^\top A\phi_h^{(0)}(u)\right)^\top$ and $\frac{\partial}{\partial h}\phi_h^{(1)}(u)=\left(\mathbf{i},\phi_h^{(0)}(u)^\top A\phi_h^{(1)}(u)\right)^\top$, respectively, with initial values $\phi_0^{(0)}(u)=\phi_0^{(1)}(u)=0$ and matrix $A:=(a_{i,j})_{i,j=1,2}\in\mathbb{R}^2$, where $a_{11}=1$, $a_{12}=a_{21}=\xi\rho$, and $a_{22}=\xi^2$. Hence, by using Ito's formula together with $\frac{d}{dt}\langle\widetilde{X},X\rangle_t=\sigma_t(1,\xi\rho)^\top$, we conclude that the L^2-optimal strategy θ is given as

$$\begin{aligned}
t\quad\mapsto\quad\theta_t &= \frac{\frac{d}{dt}\langle V^{\mathbb{Q}},X\rangle_t}{\frac{d}{dt}\langle X\rangle_t} \\
&= -\frac{1}{\sigma_t}\int_0^\infty \operatorname{Re}\left(\frac{\frac{d}{dt}\left\langle(K_{\text{str}}-I_{0,\cdot})\exp\left(-\mathbf{i}u(K_{\text{str}}-I_{0,\cdot})\right)\exp\left(\phi_{T-\cdot}^{(0)}(u)^\top\widetilde{X}\right),X\right\rangle_t}{\mathbf{i}\pi u}\right)du-\frac{\frac{d}{dt}\left\langle\frac{T-\cdot}{2T}X,X\right\rangle_t}{\sigma_t} \\
&\quad-\frac{1}{\sigma_t}\int_0^\infty \operatorname{Re}\left(\frac{\frac{d}{dt}\left\langle\exp\left(-\mathbf{i}u(K_{\text{str}}-I_{0,\cdot})\right)\exp\left(\phi_{T-\cdot}^{(0)}(u)^\top\widetilde{X}\right)\phi_{T-\cdot}^{(1)}(u)^\top\widetilde{X},X\right\rangle_t}{\pi u}\right)du \\
&= -(K_{\text{str}}-I_{0,t})\int_0^\infty \operatorname{Re}\left(\frac{\exp\left(-\mathbf{i}u(K_{\text{str}}-I_{0,t})\right)\exp\left(\phi_{T-t}^{(0)}(u)^\top\widetilde{X}_t\right)\phi_{T-t}^{(0)}(u)^\top(1,\xi\rho)^\top}{\mathbf{i}\pi u}\right)du-\frac{T-t}{2T} \\
&\quad-\int_0^\infty \operatorname{Re}\left(\frac{\exp\left(-\mathbf{i}u(K_{\text{str}}-I_{0,t})\right)\exp\left(\phi_{T-t}^{(0)}(u)^\top\widetilde{X}_t\right)\left(\phi_{T-t}^{(1)}(u)^\top\widetilde{X}_t\phi_{T-t}^{(0)}(u)+\phi_{T-t}^{(1)}(u)\right)^\top(1,\xi\rho)^\top}{\pi u}\right)du.
\end{aligned} \tag{18}$$

We then apply the fractional fast Fourier transform (FFT) in [6] to compute the integrals in (18).

For the numerical experiment in Figure 2, we choose the initial values $(X_0,\sigma_0)=(100,4)$ as well as the parameters $T=1$, $K_{\text{str}}=101$, $\kappa=0.5$, $\mu=100$, $\alpha=1$, $\beta=5$, $\xi=1$, and $\rho=-0.7$.



(a) Learning performance

(b) Payoff distribution on test set

(c) Running time and number of parameters

(d) θ and $\vartheta^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})}$ for two samples of test set

Figure 2. Learning the L^2-optimal strategy of the Asian put option G defined in (17), for $N = 0, ..., 6$. In (a), the $L^2(\mathbb{P})$-hedging error (13) and the empirical $L^2(dt \otimes \mathbb{P})$-strategy error $\left(\frac{T}{KL} \sum_{l=1}^{L} \sum_{k=0}^{K} \left| \theta_{t_k}(\omega_l) - \vartheta_{t_k}^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})}(\omega_l) \right|^2 \right)^{1/2}$ are displayed, where θ is defined in (18). In (b), the distributions of G (label "True") and $\widetilde{\varphi}_0 + \int_0^T \vartheta_t^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})} dX_t$ (label "$N = 0, ..., 6$") are depicted on the test set. In (c), the running time[4] and the number of estimated parameters are shown. In (d), we compare θ (label "True") and $\vartheta^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})}$ (label "$N = 1, ..., 6$") on two samples of the test set.

6.3. **Basket option in Wishart affine stochastic correlation model.** In the third example (see Figure 3), we consider a d-dimensional affine model with stochastic correlation, i.e. the stock price $X := (X_t)_{T \in [0,T]}$ follows the SDE

$$\begin{cases} dX_t = \kappa(\mu - X_t)dt + \sqrt{\Sigma_t} dB_t, \\ d\Sigma_t = \left(\beta U^\top U + V \Sigma_t + \Sigma_t V^\top \right) dt + \sqrt{\Sigma_t} d\widetilde{B}_t U + U^\top d\widetilde{B}_t^\top \sqrt{\Sigma_t}, \end{cases} \qquad t \in [0,T],$$

with initial values $(X_0, \Sigma_0) \in \mathbb{R}^d \times \mathbb{S}^{d-1}$ as well as parameters $\mu \in \mathbb{R}^d$, $\kappa > 0$, $\beta > d - 1$, $U \in \mathbb{R}^{d \times d}$ being invertible, and $V \in \mathbb{R}^{d \times d}$. Hereby, the d-dimensional Brownian motion $B := (B_t)_{t \in [0,T]}$ and the $(d \times d)$-dimensional Brownian motion $\widetilde{B} := \left(\widetilde{B}_t \right)_{t \in [0,T]}$ are correlated with vector $\rho \in [-1,1]^d$ satisfying $\|\rho\| \leq 1$, i.e. $B_t = \widetilde{B}_t \rho + \sqrt{1 - \rho^\top \rho} B_t^\perp$ for all $t \in [0,T]$ and a d-dimensional Brownian motion $B^\perp := \left(B_t^\perp \right)_{t \in [0,T]}$ which is independent of $\widetilde{B}$. Since $\widetilde{X} := \left(X_t, \operatorname{vec}(\Sigma_t) \right)_{t \in [0,T]}$ is an $\mathbb{R}^{d+d^2}$-valued polynomial diffusion with diffusion coefficient $a : \mathbb{R}^{d+d^2} \to \mathbb{S}_+^{d+d^2}$ of linear growth, where $\operatorname{vec}(\Sigma_t) := \left(\Sigma_t^{1,1}, \Sigma_t^{1,2}, ..., \Sigma_t^{d,d} \right)_{t \in [0,T]}^\top \in \mathbb{R}^{d^2}$, Example 2.3 ensures that $\widetilde{X}$ satisfies Assumption 2.1, which thus also applies to X. In this setting, we approximately solve the L^2-hedging problem for the Basket option

$$G = \max\left(K_{\text{str}} - w^\top X_T, 0 \right), \tag{19}$$

where $K_{\text{str}} \in \mathbb{R}$ denotes the strike price and $w \in \mathbb{R}^d$ is a fixed vector.

In order to compute the optimal strategy in the sense of quadratic hedging (see [52, 61]), we follow the arguments of the previous section. The market price of risk $t \mapsto \lambda_t := \Sigma_t^{-1} \kappa(\mu - X_t)$ induces the minimal equivalent local martingale measure $\mathbb{Q} \sim \mathbb{P}$ whose density $\frac{d\mathbb{Q}}{d\mathbb{P}} := Z_T$ is the terminal value of the SDE $dZ_t = -Z_t \lambda_t^\top \sqrt{\Sigma_t} dB_t$, with $Z_0 = 1$. Hence, the process X is a local martingale under $\mathbb{Q}$ satisfying $dX_t = \sqrt{\Sigma_t} dB_t^{\mathbb{Q}}$, where $B^{\mathbb{Q}} := \left(B_t^{\mathbb{Q}}\right)_{t\in[0,T]}$ is a d-dimensional Brownian motion under $\mathbb{Q}$. Thus, [61, Theorem 3.5] shows that $t \mapsto V_t^{\mathbb{Q}} := \mathbb{E}^{\mathbb{Q}}[G|\mathcal{F}_t]$ is the value process of the pseudo-optimal L^2-strategy in the sense of [61, p. 17]. For the calculation of the conditional expectation $\mathbb{E}^{\mathbb{Q}}[G|\mathcal{F}_t]$, we use the Fourier arguments in [4] to conclude for every $t \in [0,T]$ that

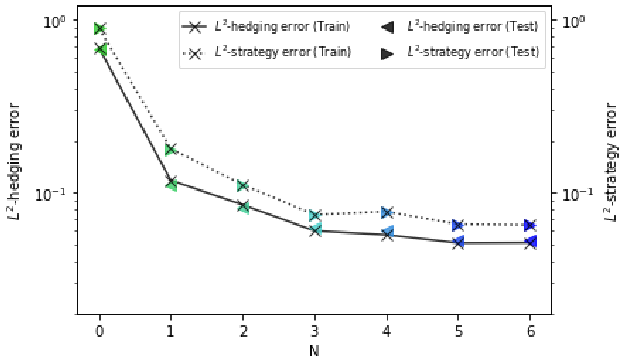
$$
\begin{aligned}
V_t^{\mathbb{Q}} &= \mathbb{E}^{\mathbb{Q}}[G|\mathcal{F}_t] = \mathbb{E}^{\mathbb{Q}}\left[\max\left(K_{\text{str}} - w^\top X_T, 0\right)\Big|\mathcal{F}_t\right] \\
&= K_{\text{str}} \mathbb{E}^{\mathbb{Q}}\left[\mathbb{1}_{\{w^\top X_T \leq K_{\text{str}}\}}\Big|\mathcal{F}_t\right] - \mathbb{E}^{\mathbb{Q}}\left[w^\top X_T \mathbb{1}_{\{w^\top X_T \leq K_{\text{str}}\}}\Big|\mathcal{F}_t\right] \\
&= K_{\text{str}}\left(\frac{1}{2} - \int_0^\infty \operatorname{Re}\left(\frac{e^{-\mathbf{i}uK}\mathbb{E}^{\mathbb{Q}}\left[\exp\left(\mathbf{i}uw^\top X_T\right)\Big|\mathcal{F}_t\right]}{\mathbf{i}\pi u}\right) du\right) \\
&\quad - \left(\frac{\mathbb{E}^{\mathbb{Q}}\left[w^\top X_T\big|\mathcal{F}_t\right]}{2} + \int_0^\infty \operatorname{Re}\left(\frac{e^{-\mathbf{i}uK}\mathbb{E}^{\mathbb{Q}}\left[\mathbf{i}w^\top X_T \exp\left(\mathbf{i}uw^\top X_T\right)\Big|\mathcal{F}_t\right]}{\pi u}\right) du\right) \\
&= K_{\text{str}}\left(\frac{1}{2} - \int_0^\infty \operatorname{Re}\left(\frac{e^{-\mathbf{i}uK}\Phi^{(0)}_{T-t,w}(u;X_t,\Sigma_t)}{\mathbf{i}\pi u}\right) du\right) \\
&\quad - \frac{w^\top X_t}{2} - \int_0^\infty \operatorname{Re}\left(\frac{e^{-\mathbf{i}uK}\Phi^{(1)}_{T-t,w}(u;X_t,\Sigma_t)}{\pi u}\right) du.
\end{aligned}
$$

Hereby, $\mathbb{R} \ni u \mapsto \Phi^{(0)}_{h,w}(u;x,s) := \mathbb{E}^{\mathbb{Q}}\left[\exp\left(\mathbf{i}uw^\top X_{t+h}\right)\big|(X_t,\Sigma_t) = (x,s)\right] \in \mathbb{C}$ denotes the conditional characteristic function of $w^\top X_{t+h}$ given that $(X_t,\Sigma_t) = (x,s) \in \mathbb{R}^d \times \mathbb{S}^d_+$, and where $\mathbb{R} \ni u \mapsto \Phi^{(1)}_{h,w}(u;x,s) := \mathbb{E}^{\mathbb{Q}}\left[\mathbf{i}w^\top X_{t+h} \exp\left(\mathbf{i}uw^\top X_{t+h}\right)\big|(X_t,\Sigma_t) = (x,s)\right] \in \mathbb{C}$ is its derivative, which can be both computed via the affine transform formula in [13] (see also the code for more details). Hence, by using Ito's formula, the notations $\nabla_x \Phi^{(r)}_{h,w}(u,x,s) := \left(\frac{\partial}{\partial x_i}\Phi^{(r)}_{h,w}(u,x,s)\right)^\top_{i=1,\dots,d}$, $\nabla_s \Phi^{(r)}_{h,w}(u,x,s) := \left(\partial^s_{i,j}\Phi^{(r)}_{h,w}(u,x,s)\right)_{i,j=1,\dots,d}$, and $\partial^s_{i,j}\Phi^{(r)}_{h,w}(u,x,s) := \frac{\partial}{\partial s_{i,j}}\Phi^{(r)}_{h,w}(u,x,s)$, for $r \in \{0,1\}$, and that $\frac{d}{dt}\langle X\rangle_t = \Sigma_t$ as well as $\frac{d}{dt}\langle \Sigma^{i,j}, X^k\rangle_t = \Sigma_t^{k,i}(U^\top\rho)_j + \Sigma_t^{k,j}(U^\top\rho)_i$, we conclude that the pseudo-optimal L^2-strategy θ is for every $t \in [0,T]$ given as

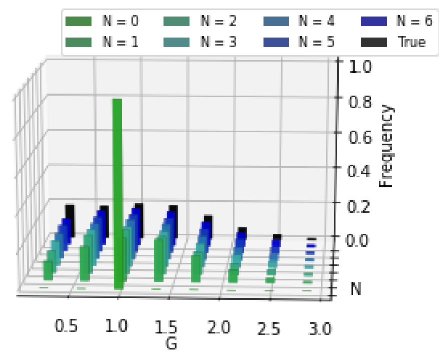
$$
\begin{aligned}
t \quad \mapsto \quad \theta_t &= \left(\frac{d}{dt}\langle X\rangle_t\right)^{-1}\frac{d}{dt}\langle V^{\mathbb{Q}}, X\rangle_t \\
&= -\Sigma_t^{-1}K_{\text{str}}\int_0^\infty \operatorname{Re}\left(\frac{e^{-\mathbf{i}uK}\frac{d}{dt}\left\langle \Phi^{(0)}_{T-\cdot,w}(u;X,\Sigma), X\right\rangle_t}{\mathbf{i}\pi u}\right) du \\
&\quad - \Sigma_t^{-1}\frac{\frac{d}{dt}\langle w^\top X, X\rangle_t}{2} - \Sigma_t^{-1}\int_0^\infty \operatorname{Re}\left(\frac{e^{-\mathbf{i}uK}\frac{d}{dt}\left\langle \Phi^{(1)}_{T-\cdot,w}(u;X,\Sigma), X\right\rangle_t}{\pi u}\right) du \\
&= -\Sigma_t^{-1}K_{\text{str}}\int_0^\infty \operatorname{Re}\left(\frac{e^{-\mathbf{i}uK}\left(\frac{d}{dt}\langle X\rangle_t \nabla_x\Phi^{(0)}_{T-t,w}(u;X_t,\Sigma_t) + \sum_{i,j=1}^d \partial^s_{i,j}\Phi^{(0)}_{T-t,w}(u;X_t,\Sigma_t)\frac{d}{dt}\langle\Sigma^{i,j},X\rangle_t\right)}{\mathbf{i}\pi u}\right) du \\
&\quad - \frac{w}{2} - \Sigma_t^{-1}\int_0^\infty \operatorname{Re}\left(\frac{e^{-\mathbf{i}uK}\left(\frac{d}{dt}\langle X\rangle_t \nabla_x\Phi^{(1)}_{T-t,w}(u;X_t,\Sigma_t) + \sum_{i,j=1}^d \partial^s_{i,j}\Phi^{(1)}_{T-t,w}(u;X_t,\Sigma_t)\frac{d}{dt}\langle\Sigma^{i,j},X\rangle_t\right)}{\pi u}\right) du \\
&= -K_{\text{str}}\int_0^\infty \operatorname{Re}\left(\frac{e^{-\mathbf{i}uK}\left(\nabla_x\Phi^{(0)}_{T-t,w}(u;X_t,\Sigma_t) + \left(\nabla_s\Phi^{(0)}_{T-t,w}(u;X_t,\Sigma_t) + \nabla_s\Phi^{(0)}_{T-t,w}(u;X_t,\Sigma_t)^\top\right)U^\top\rho\right)}{\mathbf{i}\pi u}\right) du \\
&\quad - \frac{w}{2} - \int_0^\infty \operatorname{Re}\left(\frac{e^{-\mathbf{i}uK}\left(\nabla_x\Phi^{(1)}_{T-t,w}(u;X_t,\Sigma_t) + \left(\nabla_s\Phi^{(1)}_{T-t,w}(u;X_t,\Sigma_t) + \nabla_s\Phi^{(1)}_{T-t,w}(u;X_t,\Sigma_t)^\top\right)U^\top\rho\right)}{\pi u}\right) du.
\end{aligned}
\tag{20}
$$

We then apply the fractional fast Fourier transform (FFT) in [6] to compute the integrals in (20).

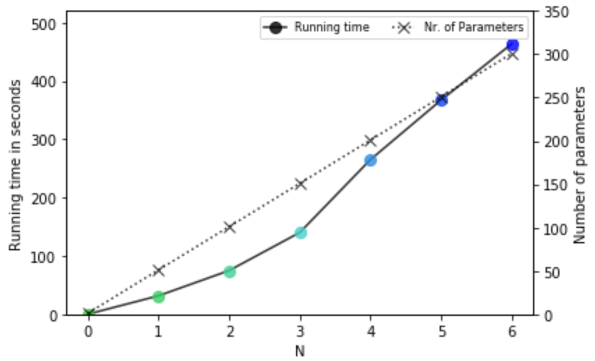
For the numerical experiment in Figure 3, we choose $d = 2$, $X_0 = \mu = (10,10)^\top$, $T = 1$, $K_{\text{str}} = 21$, $w := (1,1)^\top$, $\kappa = 1$, and $\beta = 3$, whereas the other parameters are randomly initialized.



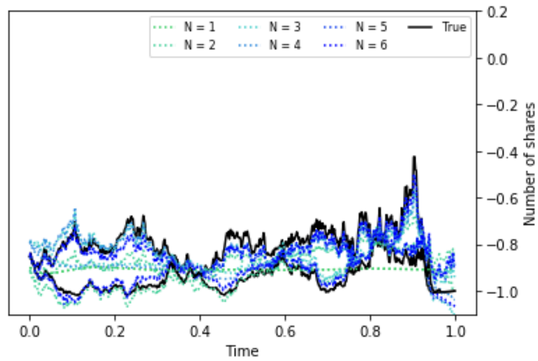
(a) Learning performance



(b) Payoff distribution on test set



(c) Running time and number of parameters



(d) θ^1 and $\vartheta^{\widetilde{\varphi}_{1:N}(\widetilde{\omega}),1}$ for two samples of test set

Figure 3. Learning the L^2-optimal strategy of the Basket option G defined in (19), for $N = 0, ..., 6$. In (a), the the empirical $L^2(\mathbb{P})$-hedging error (13) and the empirical $L^2(dt \otimes \mathbb{P})$-strategy error $\left(\frac{T}{KL} \sum_{l=1}^{L} \sum_{k=0}^{K} \left\| \theta_{t_k}(\omega_l) - \vartheta_{t_k}^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})}(\omega_l) \right\|^2 \right)^{1/2}$ are displayed, where θ is defined in (20). In (b), the distributions of G (label "True") and $\widetilde{\varphi}_0 + \int_0^T \left(\vartheta_t^{\widetilde{\varphi}_{1:N}(\widetilde{\omega})} \right)^\top dX_t$ (label "$N = 0, ..., 6$") are depicted on the test set. In (c), the running time[4] and the number of estimated parameters are shown. In (d), we compare θ^1 (label "True") and $\vartheta^{\widetilde{\varphi}_{1:N}(\widetilde{\omega}),1}$ (label "$N = 1, ..., 6$") on two samples of the test set.

7. Proofs

7.1. Proof of Results in Section 2.

Proof of Example 2.3. We adapt the proof of [21, Theorem 3.3] and assume that $X := (X_t)_{t \in [0,T]}$ is an E-valued strong solution of the SDE (3), where $E \subseteq \mathbb{R}^d$, where $E \ni x \mapsto b(x) := \beta_0 + \beta_1 x \in \mathbb{R}^d$ is the drift coefficient (with $\beta_0 \in \mathbb{R}^d$ and $\beta_1 \in \mathbb{R}^{d \times d}$), where $\sigma : E \to \mathbb{R}^{d \times d}$ is the volatility, and where $E \ni x \mapsto a(x) := \sigma(x)\sigma(x)^\top \in \mathbb{S}_+^d$ is the diffusion coefficient. Then, by using variation of constants, we conclude that $X_t = K_t + L_t$ for all $t \in [0,T]$, where $K_t := e^{\beta_1 t} X_0 + \int_0^t e^{\beta_1(t-s)} \beta_0 ds$ and $L_t := e^{\beta_1(T-t)} Y_t$, with $Y_t := \int_0^t \sigma_s^Y dB_s$ and $\sigma_s^Y := e^{\beta_1(T-s)} \sigma\big(K_s + e^{-\beta_1(T-s)} Y_s\big)$. Moreover, by using the constant $C_2 := \sup_{t \in [0,T]} \left\| e^{\beta_1(T-t)} \right\|_F > 0$ together with $\|\sigma(x)\|_F^2 = \operatorname{tr}\left(\sigma(x)\sigma(x)^\top\right) = \operatorname{tr}(a(x)) \leq \sqrt{d} \, \|a(x)\|_F$ for any $x \in E$, the linear growth condition in (4) (with constant $C_1 > 0$), and the constant $C_3 := C_1 C_2^2 \sqrt{d} \sup_{t \in [0,T]} \left(1 + \|K_t\| + \left\| e^{-\beta_1(T-t)} \right\|_F \right) > 0$, it follows for every $t \in [0,T]$ that

$$
\begin{aligned}
\left\| \sigma_t^Y \right\|_F^2 &= \left\| e^{\beta_1(T-t)} \sigma\left(K_t + e^{-\beta_1(T-t)} Y_t\right) \right\|_F^2 \leq \left\| e^{\beta_1(T-t)} \right\|_F^2 \left\| \sigma\left(K_t + e^{-\beta_1(T-t)} Y_t\right) \right\|_F^2 \\
&\leq C_2^2 \sqrt{d} \left\| a\left(K_t + e^{-\beta_1(T-t)} Y_t\right) \right\|_F \leq C_1 C_2^2 \sqrt{d} \left(1 + \left\| K_t + e^{-\beta_1(T-t)} Y_t \right\| \right) \\
&\leq C_1 C_2^2 \sqrt{d} \left(1 + \|K_t\| + \left\| e^{-\beta_1(T-t)} \right\|_F \|Y_t\| \right) \leq C_3 \left(1 + \|Y_t\|\right).
\end{aligned} \tag{21}
$$

Now, let $f : \mathbb{R}^d \to \mathbb{R}$ be a smooth function such that $\sqrt{\|y\|} \leq f(y)$ for all $y \in \mathbb{R}^d$ with $\|y\| \leq 1$ as well as $f(y) = \sqrt{1+\|y\|}$ for all $y \in \mathbb{R}^d$ with $\|y\| > 1$. Then, the first and second order partial derivatives $\nabla f := \left(\frac{\partial f}{\partial y_i}\right)_{i=1,\dots,d}^\top : \mathbb{R}^d \to \mathbb{R}^d$ and $\nabla^2 f := \left(\frac{\partial^2 f}{\partial y_i \partial y_j}\right)_{i,j=1,\dots,d} : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ can be computed on $\{y \in \mathbb{R}^d : \|y\| > 1\}$ (see [21, p. 954]). Moreover, by applying Ito's formula on the process $t \mapsto Z_t := f(Y_t)$, we conclude that $dZ_t = \mu_t^Z dt + \sigma_t^Z dB_t$, $\mathbb{P}$-a.s., for all $t \in [0,T]$, where $\mu_t^Z := \frac{1}{2} \operatorname{tr}\left(\nabla^2 f(Y_t) \sigma_t^Y (\sigma_t^Y)^\top\right)$ and $\sigma_t^Z := \nabla f(Y_t)^\top \sigma_t^Y dB_t$. Hence, by using that $\nabla f : \mathbb{R}^d \to \mathbb{R}^d$ and $\nabla^2 f : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are uniformly bounded on $\mathbb{R}^d$ together with (21), there exists some $m, \rho > 0$ such that for every $t \in [0,T]$ it holds that $\mu_t^Z \leq m$ and $\left\|\sigma_t^Z\right\| \leq \rho$. Thus, for $\delta \in (0, 1/(2\rho^2 T))$, we can apply [27, Theorem 1.3] with the convex and non-decreasing function $\mathbb{R} \ni s \mapsto e^{\delta s^2} \in \mathbb{R}$ to conclude for every $t \in [0,T]$ and a corresponding normal distributed random variable $V_t \sim \mathcal{N}(f(0)+mt, \rho^2 t)$ that

$$\mathbb{E}\left[e^{\delta Z_t^2}\right] \leq \mathbb{E}\left[e^{\delta V_t^2}\right] \leq \mathbb{E}\left[e^{\delta V_T^2}\right] < \infty. \tag{22}$$

Next, we fix some $\varepsilon > 0$ with $C_2\varepsilon \leq \delta$ and $C_1 \varepsilon T \leq \delta$. Then, by using the triangle inequality, the constants $C_4 := \sup_{t\in[0,T]} \|K_t\| > 0$ and $C_2 > 0$ (defined above), that $\|Y_t\| \leq f(Y_t)^2 = Z_t^2$ for any $t \in [0,T]$, and the inequality (22), it follows for every $t \in [0,T]$ that

$$\mathbb{E}\left[e^{\varepsilon\|X_t\|}\right] \leq \mathbb{E}\left[e^{\varepsilon(\|K_t\|+\|L_t\|)}\right] \leq e^{C_4\varepsilon} \mathbb{E}\left[e^{C_2\varepsilon\|Y_t\|}\right] \leq e^{C_4\varepsilon}\mathbb{E}\left[e^{\delta Z_t^2}\right] \leq e^{C_4\varepsilon}\mathbb{E}\left[e^{\delta V_T^2}\right] < \infty, \tag{23}$$

which shows that X satisfies Assumption 2.1 (i). Moreover, by using that the finite variation part A satisfies $A_t = \int_0^t b(X_s) ds$ for any $t \in [0,T]$ together with Jensen's inequality and Fubini's theorem, Minkowski's inequality together with Jensen's inequality and $n := \lceil \gamma \rceil \in \mathbb{N}_0$, that $\frac{x^n}{n!} \leq \exp(x)$ for any $n \in \mathbb{N}_0$ and $x \in [0,\infty)$, and the inequality (23), it holds for every $\gamma \in [1,\infty)$ that

$$\begin{aligned}
\mathbb{E}\left[\left(\int_0^T \|dA_t\|\right)^\gamma\right]^{\frac{1}{\gamma}} &\leq T^{1-\frac{1}{\gamma}} \left(\int_0^T \mathbb{E}\left[\|b(X_t)\|^\gamma\right] dt\right)^{\frac{1}{\gamma}} \leq T \sup_{t\in[0,T]} \mathbb{E}\left[(\|\beta_0\| + \|\beta_1\|_F \|X_t\|)^\gamma\right]^{\frac{1}{\gamma}} \\
&\leq \|\beta_0\| T + \|\beta_1\|_F T \sup_{t\in[0,T]} \mathbb{E}\left[\|X_t\|^n\right]^{\frac{1}{n}} = \|\beta_0\|T + \|\beta_1\|_F T \frac{(n!)^{\frac{1}{n}}}{\varepsilon} \sup_{t\in[0,T]} \mathbb{E}\left[\frac{(\varepsilon\|X_t\|)^n}{n!}\right]^{\frac{1}{n}} \\
&\leq \|\beta_0\|T + \|\beta_1\|_F T \frac{(n!)^{\frac{1}{n}}}{\varepsilon} \sup_{t\in[0,T]} \mathbb{E}\left[e^{\varepsilon\|X_t\|}\right]^{\frac{1}{n}} \leq \|\beta_0\|T + \|\beta_1\|_F T \frac{(n!)^{\frac{1}{n}}}{\varepsilon} e^{C_4 \frac{\varepsilon}{n}} \mathbb{E}\left[e^{\delta V_T^2}\right]^{\frac{1}{n}} < \infty,
\end{aligned}$$

which shows that X satisfies Assumption 2.1 (ii). In addition, by using that the local martingale part M satisfies $M_t = \int_0^t \sigma(X_s) dB_s$ for any $t \in [0,T]$ together with Jensen's inequality, the linear growth condition in (4) (with constant $C_1 > 0$), Minkowski's inequality together with Jensen's inequality and $n := \lceil \gamma \rceil \in \mathbb{N}_0$, and the inequality (23), we conclude for every $\gamma \in [1,\infty)$ that

$$\begin{aligned}
\mathbb{E}\left[\|\langle M\rangle_T\|_F^{\gamma}\right]^{\frac{1}{c}} &\leq T^{1-\frac{1}{\gamma}} \left(\int_0^T \mathbb{E}\left[\|a(X_t)\|_F^\gamma\right] dt\right)^{\frac{1}{\gamma}} \leq C_1 T \sup_{t\in[0,T]} \mathbb{E}\left[(1+\|X_t\|)^\gamma\right]^{\frac{1}{\gamma}} \\
&\leq C_1 T + C_1 T \sup_{t\in[0,T]} \mathbb{E}\left[\|X_t\|^n\right]^{\frac{1}{n}} = C_1 T + C_1 T \frac{(n!)^{\frac{1}{n}}}{\varepsilon} \sup_{t\in[0,T]} \mathbb{E}\left[\frac{(\varepsilon\|X_t\|)^n}{n!}\right]^{\frac{1}{n}} \\
&\leq C_1 T + C_1 T \frac{(n!)^{\frac{1}{n}}}{\varepsilon} \sup_{t\in[0,T]} \mathbb{E}\left[e^{\varepsilon\|X_t\|}\right] \leq C_1 T + C_1 T \frac{(n!)^{\frac{1}{n}}}{\varepsilon} e^{C_4\frac{\varepsilon}{n}} \mathbb{E}\left[e^{\delta V_T^2}\right]^{\frac{1}{n}} < \infty,
\end{aligned}$$

which shows that X satisfies Assumption 2.1 (iii). ☐

7.2. **Proof of results in Section 3.**

7.2.1. *Proof of results in Section 3.1.*

Proof of Lemma 3.2. For (i), we first use that the space of càdlàg functions forms a vector space to conclude that we only need to show that $\|\cdot\|_{L^p(X)}$ is a norm. By definition, $\|\cdot\|_{L^p(X)}$ is absolutely homogeneous. Moreover, $\|\cdot\|_{L^p(X)}$ is positive definite as $L^p(X)$ is defined in terms of equivalence classes. In order to show the triangle inequality, let $f, g \in L^p(X)$. Then, by using the Kunita-Watanabe inequality to the continuous local martingales $t \mapsto M_t^f := \int_0^t f(s)^\top dM_s$ and $t \mapsto M_t^g := \int_0^t g(s)^\top dM_s$, and the identity $x + 2\sqrt{xy} + y = (\sqrt{x} + \sqrt{y})^2$ for any $x, y \geq 0$, it holds $\mathbb{P}$-a.s. that

$$
\begin{aligned}
&\int_0^T (f(t)+g(t))^\top d\langle M\rangle_t (f(t)+g(t)) \\
&\quad = \int_0^T f(t)^\top d\langle M\rangle_t f(t) + 2 \underbrace{\int_0^T f(t)^\top d\langle M\rangle_t g(t)}_{=\int_0^T d\langle M^f, M^g\rangle_t} + \int_0^T g(t)^\top d\langle M\rangle_t g(t) \\
&\quad \leq \langle M^f\rangle_T + 2\langle M^f\rangle_T^{\frac{1}{2}} \langle M^g\rangle_T^{\frac{1}{2}} + \langle M^g\rangle_T = \left(\langle M^f\rangle_T^{\frac{1}{2}} + \langle M^g\rangle_T^{\frac{1}{2}} \right)^2.
\end{aligned}
$$

Finally, by using the triangle inequality and Minkowski's inequality, we conclude that

$$
\begin{aligned}
&\|f+g\|_{L^p(X)} = \mathbb{E}\left[\left(\int_0^T \left| (f(t)+g(t))^\top dA_t \right| \right)^p \right]^{\frac{1}{p}} + \mathbb{E}\left[\left(\int_0^T (f(t)+g(t))^\top d\langle M\rangle_t (f(t)+g(t)) \right)^{\frac{p}{2}} \right]^{\frac{1}{p}} \\
&\leq \mathbb{E}\left[\left(\int_0^T \left| f(t)^\top dA_t \right| + \int_0^T \left| g(t)^\top dA_t \right| \right)^p \right]^{\frac{1}{p}} + \mathbb{E}\left[\left(\langle M^f\rangle_T^{\frac{1}{2}} + \langle M^g\rangle_T^{\frac{1}{2}} \right)^p \right]^{\frac{1}{p}} \\
&\leq \mathbb{E}\left[\left(\int_0^T \left| f(t)^\top dA_t \right| \right)^p \right]^{\frac{1}{p}} + \mathbb{E}\left[\left(\int_0^T \left| g(t)^\top dA_t \right| \right)^p \right]^{\frac{1}{p}} + \mathbb{E}\left[\langle M^f\rangle_T^{\frac{p}{2}} \right]^{\frac{1}{p}} + \mathbb{E}\left[\langle M^g\rangle_T^{\frac{p}{2}} \right]^{\frac{1}{p}} \\
&= \|f\|_{L^p(X)} + \|g\|_{L^p(X)},
\end{aligned}
$$

which shows (i).

For (ii), we first recall that every fixed $g := (g_1, ..., g_d)^\top \in L^p(X)$ is $\mathcal{B}([0,T])/\mathcal{B}(\mathbb{R}^d)$-measurable and bounded (see Remark 3.1), i.e. $\|g\|_\infty := \sup_{t \in [0,T]} \|g(t)\| < \infty$. Moreover, by using that X satisfies Assumption 2.1, we can define the finite constant

$$
C_{p,T,X} := \mathbb{E}\left[\left(\int_0^T \|dA_t\| \right)^p \right]^{\frac{1}{p}} + \sqrt{3d}\,\mathbb{E}\left[\|\langle M\rangle_T\|_F^{\frac{p}{2}} \right]^{\frac{1}{p}} < \infty. \tag{24}
$$

In addition, by using the polarization identity $\langle M^i, M^j\rangle_t = \frac{1}{2}\left(\langle M^i + M^j\rangle_t - \langle M^i\rangle_t - \langle M^j\rangle_t \right)$ and the inequality $\sum_{i,j=1}^d c_{i,j} \leq d\left(\sum_{i,j=1}^d c_{i,j}^2 \right)^{1/2}$ for any $c_{i,j} \geq 0$, we conclude that

$$
\begin{aligned}
&\int_0^T g(t)^\top d\langle M\rangle_t g(t) = \frac{1}{2} \left| \sum_{i,j=1}^d \int_0^T g_i(t) \left(d\langle M^i + M^j\rangle_t - d\langle M^i\rangle_t - d\langle M^j\rangle_t \right) g_j(t) \right| \\
&\leq \frac{1}{2} \sum_{i,j=1}^d \left(\left| \int_0^T g_i(t) d\langle M^i + M^j\rangle_t g_j(t) \right| + \left| \int_0^T g_i(t) d\langle M^i\rangle_t g_j(t) \right| + \left| \int_0^T g_i(t) d\langle M^j\rangle_t g_j(t) \right| \right) \\
&\leq \frac{\|g\|_\infty^2}{2} \sum_{i,j=1}^d \left(\langle M^i + M^j\rangle_T + \langle M^i\rangle_T + \langle M^j\rangle_T \right) \leq 3\|g\|_\infty^2 \sum_{i,j=1}^d |\langle M^i, M^j\rangle_T| \leq 3d\|g\|_\infty^2 \|\langle M\rangle_T\|_F.
\end{aligned} \tag{25}
$$

Hence, by using the Cauchy-Schwarz inequality and the inequalities (24)+(25), we obtain

$$
\begin{aligned}
\|g\|_{L^p(X)} &= \mathbb{E}\left[\left(\int_0^T \left| g(t)^\top dA_t \right| \right)^p \right]^{\frac{1}{p}} + \mathbb{E}\left[\left(\int_0^T g(t)^\top d\langle M\rangle_t g(t) \right)^{\frac{p}{2}} \right]^{\frac{1}{p}} \\
&\leq \|g\|_\infty \mathbb{E}\left[\left(\int_0^T \|dA_t\| \right)^p \right]^{\frac{1}{p}} + \sqrt{3d}\|g\|_\infty \mathbb{E}\left[\|\langle M\rangle_T\|_F^{\frac{p}{2}} \right]^{\frac{1}{p}} \leq C_{p,T,X} \|g\|_\infty < \infty,
\end{aligned}
$$

which completes the proof. □

Proof of Lemma 3.5. Fix some $g \in L^p(X)$. Then, by using Minkowski's inequality and the Burkholder-Davis-Gundy inequality (with constant $C_p > 0$), it follows that

$$\begin{aligned}\mathbb{E}\left[\sup_{t\in[0,T]}|W(g)_t|^p\right]^{\frac{1}{p}} &\leq \mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t g(s)^\top dA_s\right|^p\right]^{\frac{1}{p}} + \mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t g(s)^\top dM_s\right|^p\right]^{\frac{1}{p}} \\ &\leq \mathbb{E}\left[\left(\int_0^T \left|g(s)^\top dA_s\right|\right)^p\right]^{\frac{1}{p}} + C_p\mathbb{E}\left[\left(\int_0^T g(s)^\top d\langle M\rangle_s g(s)\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} \\ &\leq \max(1, C_p)\,\|g\|_{L^p(X)},\end{aligned}$$

which completes the proof. □

7.2.2. *Proof of results in Section 3.2.*

Proof of Lemma 3.6. First, we consider two Banach spaces $(Y, \|\cdot\|_Y)$ and $(Z, \|\cdot\|_Z)$. Then, by using [59, Proposition 2.1], the tensor product $(Y \otimes Z, \|\cdot\|_{Y\otimes Z})$ is again a Banach space with projective norm

$$\|x\|_{Y\otimes Z} = \inf\left\{\sum_{j=1}^m \|y_j\|_Y\|z_j\|_Z : x = \sum_{j=1}^m y_j \otimes z_j,\ m\in\mathbb{N},\ y_1,...,y_m \in Y,\ z_1,...,z_m \in Z\right\},$$

where it holds that $\|y\otimes z\|_{Y\otimes Z} = \|y\|_Y\|z\|_Z$ for all $y \in Y$ and $z \in Z$. Moreover, by following the proof of [59, Proposition 2.1], we observe that the completeness of $(Y, \|\cdot\|_Y)$ and $(Z, \|\cdot\|_Z)$ has never been used. Hence, if $(Y, \|\cdot\|_Y)$ and $(Z, \|\cdot\|_Z)$ are normed vector space, then $(Y \otimes Z, \|\cdot\|_{Y\otimes Z})$ is a normed vector space with $\|y\otimes z\|_{Y\otimes Z} = \|y\|_Y\|z\|_Z$ for all $y \in Y$ and $z \in Z$.

Thus, by using the above arguments iteratively, we can show that $(L^{np}(X)^{\otimes n}, \|\cdot\|_{L^{np}(X)^{\otimes n}})$ is a normed vector space such that for every $g_1, ..., g_n \in L^{np}(X)$ it holds that

$$\|g_1\otimes\cdots\otimes g_n\|_{L^{np}(X)^{\otimes n}} = \|g_1\otimes\cdots\otimes g_{n-1}\|_{L^{np}(X)^{\otimes(n-1)}}\|g_n\|_{L^{np}(X)} = ... = \prod_{k=1}^n \|g_k\|_{L^{np}(X)},$$

which completes the proof. □

Proof of Lemma 3.9. For (i), we first define $L^{np}(X)\times\cdots\times L^{np}(X) \ni (g_1, ..., g_n) \mapsto L(g_1, ..., g_n) := J_n^\circ(g_1\otimes\cdots\otimes g_n)_t \to L^0(\Omega, \mathcal{F}_t, \mathbb{P})$, which is by (7) multilinear, where $L^0(\Omega, \mathcal{F}_t, \mathbb{P})$ denotes the space of $\mathcal{F}_t$-measurable random variables. Moreover, by using the universal property of the tensor product in [59, Proposition 1.4], the map

$$L^{np}(X)^{\otimes n} \ni g = \sum_{j=1}^m g_{j,1}\otimes\cdots\otimes g_{j,n} \quad\mapsto\quad \widetilde{L}(g) := \sum_{j=1}^m L(g_{j,1}, ..., g_{j,n}) \in \mathbb{R}$$

is linear. However, since $J_n^\circ(g)_t = \sum_{j=1}^m J_n^\circ(g_{j,1}\otimes\cdots\otimes g_{j,n})_t = \sum_{j=1}^m L(g_{j,1}, ..., g_{j,n}) = \widetilde{L}(g)$ for all $g = \sum_{j=1}^m g_{j,1}\otimes\cdots\otimes g_{j,n} \in L^{np}(X)^{\otimes n}$, we conclude that $J_n^\circ(\cdot)_t = \widetilde{L}$ is linear on $L^{np}(X)^{\otimes n}$.

For (ii), we use the linearity of $J_n^\circ(\cdot)_t : L^{np}(X)^{\otimes n} \to L^0(\Omega, \mathcal{F}_t, \mathbb{P})$ in (i) to conclude that it suffices to show that $J_n^\circ(g)_t = J_n^\circ(0)_t = 0$, $\mathbb{P}$-a.s., for each representation $g = \sum_{j=1}^m g_{k,1}\otimes\cdots\otimes g_{j,n}$ of $0 \in L^{np}(X)^{\otimes n}$. For every fixed $E \in \mathcal{A}$, we define the multilinear form $L^{np}(X)\times\cdots\times L^{np}(X) \ni (f_1, ..., f_n) \mapsto L(f_1, ..., f_n) := \mathbb{E}\left[\mathbb{1}_E J_n^\circ(f_1\otimes\cdots\otimes f_n)_t\right] \in \mathbb{R}$. Then, it follows for every representation $g := \sum_{j=1}^m g_{j,1}\otimes\cdots\otimes g_{j,n} \in L^{np}(X)^{\otimes n}$ of $0 \in L^{np}(X)^{\otimes n}$ that

$$\mathbb{E}\left[\mathbb{1}_E J_n^\circ(g)_t\right] = \sum_{j=1}^m \mathbb{E}\left[\mathbb{1}_E J_n^\circ(g_{j,1}\otimes\cdots\otimes g_{j,n})_t\right] = \sum_{j=1}^m L(g_{j,1}, ..., g_{j,n}) = L(g).$$

Since the value of $L(g)$ does not dependent on the representation of $0 \in L^{np}(X)^{\otimes n}$ (see [59, p. 2]), we can choose $g = 0 \in L^{np}(X)^{\otimes n}$ to conclude that $\mathbb{E}\left[\mathbb{1}_E J_n^\circ(g)_t\right] = L(g) = 0$. However, as $E \in \mathcal{A}$ was chosen arbitrarily, we obtain $J_n^\circ(g)_t = 0$, $\mathbb{P}$-a.s., which completes the proof.

For (iii), we use induction on $n \in \mathbb{N}$. For $n = 1$, we observe that $J_1^\circ(g)_t = W(g)_t$ for all $g \in L^p(X)$ and $t \in [0, T]$, whence the conclusion follows from Lemma 3.5. Now, we fix some $n \in \mathbb{N}\cap[2,\infty)$ and $g_1, ..., g_n \in L^{np}(X)$, assume that (iii) holds true for $1, ..., n-1$, and aim to show (iii) for n. To this end, we use the notation $g_{1:(n-1)} := g_1 \otimes \cdots \otimes g_{n-1} \in L^{np}(X)^{\otimes(n-1)}$ and Hölder's inequality to conclude that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t J_{n-1}^{\circ}(g_{1:(n-1)})_s g_n(s)^\top dA_s\right|^p\right]^{\frac{1}{p}} \\
&\quad\leq \mathbb{E}\left[\sup_{t\in[0,T]}\left|J_{n-1}^{\circ}(g_{1:(n-1)})_t\right|^p\left(\int_0^T\left|g_k(s)^\top dA_s\right|\right)^p\right]^{\frac{1}{p}} \\
&\quad\leq \mathbb{E}\left[\sup_{t\in[0,T]}\left|J_{n-1}^{\circ}(g_{1:(n-1)})_t\right|^{\frac{np}{n-1}}\right]^{\frac{n-1}{np}}\mathbb{E}\left[\left(\int_0^T\left|g_k(s)^\top dA_s\right|\right)^{np}\right]^{\frac{1}{np}}.
\end{aligned} \tag{26}$$

Moreover, the Burkholder-Davis-Gundy inequality (with constant $C_p > 0$) and Hölder's inequality imply that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t J_{n-1}^{\circ}(g_{1:(n-1)})_s g_k(s)^\top dM_s\right|^p\right]^{\frac{1}{p}} \\
&\quad\leq C_p\mathbb{E}\left[\left|\int_0^T J_{n-1}^{\circ}(g_{1:(n-1)})_s^2 g_k(s)^\top d\langle M\rangle_s g_k(s)\right|^{\frac{p}{2}}\right]^{\frac{1}{p}} \\
&\quad\leq C_p\mathbb{E}\left[\sup_{t\in[0,T]}\left|J_{n-1}^{\circ}(g_{1:(n-1)})_t\right|^p\left(\int_0^T g_k(s)^\top d\langle M\rangle_s g_k(s)\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} \\
&\quad\leq C_p\mathbb{E}\left[\sup_{t\in[0,T]}\left|J_{n-1}^{\circ}(g_{1:(n-1)})_t\right|^{\frac{np}{n-1}}\right]^{\frac{n-1}{np}}\mathbb{E}\left[\left(\int_0^T g_k(s)^\top d\langle M\rangle_s g_k(s)\right)^{\frac{np}{2}}\right]^{\frac{1}{np}}.
\end{aligned} \tag{27}$$

In addition, by using the notation $g_{1:(n-2)} := g_1 \otimes \cdots \otimes g_{n-2} \in L^{np}(X)^{\otimes(n-2)}$, the Kunita-Watanabe inequality, and Hölder's inequality, we conclude that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t J_{n-2}^{\circ}(g_{1:(n-2)})_s g_{n-1}(s)^\top d\langle X\rangle_s g_n(s)\right|^p\right]^{\frac{1}{p}} \\
&\quad\leq \mathbb{E}\left[\sup_{t\in[0,T]}\left|J_{n-2}^{\circ}(g_{1:(n-2)})_t\right|^p\left|\int_0^T g_{n-1}(s)^\top d\langle M\rangle_s g_n(s)\right|^p\right]^{\frac{1}{p}} \\
&\quad\leq \mathbb{E}\left[\sup_{t\in[0,T]}\left|J_{n-2}^{\circ}(g_{1:(n-2)})_t\right|^p\left(\int_0^T g_{n-1}(s)^\top d\langle M\rangle_s g_{n-1}(s)\right)^{\frac{p}{2}}\left(\int_0^T g_n(s)^\top d\langle M\rangle_s g_n(s)\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} \\
&\quad\leq \mathbb{E}\left[\sup_{t\in[0,T]}\left|J_{n-2}^{\circ}(g_{1:(n-2)})_t\right|^{\frac{np}{n-2}}\right]^{\frac{n-2}{np}}\mathbb{E}\left[\left(\int_0^T g_{n-1}(s)^\top d\langle M\rangle_s g_{n-1}(s)\right)^{\frac{np}{2}}\right]^{\frac{1}{np}} \\
&\qquad\cdot\mathbb{E}\left[\left(\int_0^T g_n(s)^\top d\langle M\rangle_s g_n(s)\right)^{\frac{np}{2}}\right]^{\frac{1}{np}}.
\end{aligned} \tag{28}$$

Hence, by using the notation $g_{1:n} := g_1 \otimes \cdots \otimes g_n \in L^{np}(X)^{\otimes n}$, the identity (7) together with $dX_t = dM_t + dA_t$, $t \in [0,T]$, and Minkowski's inequality, the inequalities (26)-(28), and the induction hypothesis (i.e. that (iii) holds true for $1, ..., n-1$ with exponent $\frac{np}{n-1}$) together with $\|g_{1:n_1}\|_{L^{np}(X)^{\otimes n_1}} =$

$\prod_{k=1}^{n_1} \|g_k\|_{L^{np}(X)}$ for $n_1 \in \{n-2, n-1\}$ (see Lemma 3.6), we obtain that

$$
\begin{aligned}
\mathbb{E}\left[\sup_{t\in[0,T]} |J_n^\circ(g_{1:n})_t|^p\right]^{\frac{1}{p}} &\leq \mathbb{E}\left[\sup_{t\in[0,T]} \left|\int_0^t J_{n-1}^\circ(g_{1:(n-1)})_s g_n(s)^\top dA_s\right|^p\right]^{\frac{1}{p}} \\
&\quad + \mathbb{E}\left[\sup_{t\in[0,T]} \left|\int_0^t J_{n-1}^\circ(g_{1:(n-1)})_s g_n(s)^\top dM_s\right|^p\right]^{\frac{1}{p}} \\
&\quad + \frac{1}{2}\mathbb{E}\left[\sup_{t\in[0,T]} \left|\int_0^t J_{n-2}^\circ(g_{1:(n-2)})_s g_{n-1}(s)^\top d\langle X\rangle_s g_n(s)\right|^p\right]^{\frac{1}{p}} \\
&\leq \mathbb{E}\left[\sup_{t\in[0,T]} \left|J_{n-1}^\circ(g_{1:(n-1)})_t\right|^{\frac{np}{n-1}}\right]^{\frac{n-1}{np}} \mathbb{E}\left[\left(\int_0^T \left|g_k(s)^\top dA_s\right|\right)^{np}\right]^{\frac{1}{np}} \\
&\quad + C_p\mathbb{E}\left[\sup_{t\in[0,T]} \left|J_{n-1}^\circ(g_{1:(n-1)})_t\right|^{\frac{np}{n-1}}\right]^{\frac{n-1}{np}} \mathbb{E}\left[\left(\int_0^T g_k(s)^\top d\langle M\rangle_s g_k(s)\right)^{\frac{np}{2}}\right]^{\frac{1}{np}} \\
&\quad + \frac{1}{2}\mathbb{E}\left[\sup_{t\in[0,T]} \left|J_{n-2}^\circ(g_{1:(n-2)})_t\right|^{\frac{np}{n-2}}\right]^{\frac{n-2}{np}} \mathbb{E}\left[\left(\int_0^T g_{n-1}(s)^\top d\langle M\rangle_s g_{n-1}(s)\right)^{\frac{np}{2}}\right]^{\frac{1}{np}} \\
&\qquad \cdot \mathbb{E}\left[\left(\int_0^T g_n(s)^\top d\langle M\rangle_s g_n(s)\right)^{\frac{np}{2}}\right]^{\frac{1}{np}} \\
&\leq \left(\prod_{k=1}^{n-1} \max\left(1, C_{\frac{(n-1)\frac{np}{n-1}}{k}}\right)\right)\left(\prod_{k=1}^{n-1} \|g_k\|_{L^{np}(X)}\right) \max(1, C_p)\|g_n\|_{L^{np}(X)} \\
&\quad + \frac{1}{2}\left(\prod_{k=1}^{n-2} \|g_k\|_{L^{np}(X)}\right) \|g_{n-1}\|_{L^{np}(X)}\|g_n\|_{L^{np}(X)} \\
&\leq \frac{3}{2}\left(\prod_{k=1}^{n} \max\left(1, C_{\frac{np}{k}}\right)\right) \|g_1 \otimes \cdots \otimes g_n\|_{L^{np}(X)^{\otimes n}}.
\end{aligned}
\tag{29}
$$

For the general case of $g \in L^{np}(X)^{\otimes n}$, we fix some $\varepsilon > 0$ and assume that $\sum_{j=1}^m g_{j,1} \otimes \cdots \otimes g_{j,n}$ is a representation of $g \in L^{np}(X)^{\otimes n}$ such that $\sum_{j=1}^m \prod_{k=1}^n \|g_{j,k}\|_{L^{np}(X)} \leq \|g\|_{L^{np}(X)^{\otimes n}} + \varepsilon/C_{n,p}$. Then, by using Minkowski's inequality and (29), it follows that

$$
\begin{aligned}
\mathbb{E}\left[\sup_{t\in[0,T]} |J_n^\circ(g)_t|^p\right]^{\frac{1}{p}} &\leq \sum_{j=1}^m \mathbb{E}\left[\sup_{t\in[0,T]} |J_n^\circ(g_{j,1} \otimes \cdots \otimes g_{j,n})_t|^p\right]^{\frac{1}{p}} \\
&\leq C_{n,p} \sum_{j=1}^m \prod_{k=1}^n \|g_{j,k}\|_{L^{np}(X)} \\
&\leq C_{n,p}\|g\|_{L^{np}(X)^{\otimes n}} + \varepsilon.
\end{aligned}
$$

Since $\varepsilon > 0$ was chosen arbitrarily, we obtain the conclusion in (iii). □

Proof of Proposition 3.10. Fix some $g = \sum_{j=1}^m g_{j,1} \otimes \cdots \otimes g_{j,n} \in L^{np}_{\mathrm{sym}}(X)^{\otimes n}$ and $t \in [0,T]$. Then, we show the conclusion by induction over $n \in \mathbb{N}$. For $n = 1$, we observe that

$$
J_1^\circ(g)_t = \sum_{j=1}^m \int_0^t \circ dW(g_{j,1})_t = \sum_{j=1}^m W(g_{j,1})_t.
$$

Now, we fix some $n \in \mathbb{N} \cap [2,\infty)$, assume that the conclusion holds true for $n-1$, and aim to show that it also holds true for n. Indeed, by using that $g = \mathrm{sym}(g)$ together with the linearity of $J_n^\circ : L^{np}(X)^{\otimes n} \to L^p(\Omega, \mathcal{F}_t, \mathbb{P})$, the set $\mathcal{S}_{n,k}$ consisting of all permutations $\sigma : \{1,...,n\} \setminus \{k\} \to$

$\{1,...,n\} \setminus \{k\}$, $l = 1,...,n$, the induction hypothesis (i.e. that the conclusion holds true for $n-1$), and the definition of the Stratonovich integral, it follows that

$$
\begin{aligned}
J_n^\circ(g) = J_n^\circ(\mathrm{sym}(g)) &= \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} \sum_{j=1}^m J_n^\circ(g_{j,\sigma(1)} \otimes \cdots \otimes g_{j,\sigma(n)}) \\
&= \frac{1}{n!} \sum_{j=1}^m \sum_{k=1}^n \sum_{\sigma \in \mathcal{S}_{n,k}} \int_0^t J_{n-1}^\circ \left(g_{j,\sigma(1)} \otimes \cdots \otimes g_{j,\sigma(k)-1} \otimes g_{j,\sigma(k)+1} \otimes \cdots \otimes g_{j,\sigma(n)} \right)_s \circ dW(g_{j,\sigma(k)})_s \\
&= \frac{1}{n} \sum_{j=1}^m \sum_{k=1}^n \int_0^t J_{n-1}^\circ \left(\mathrm{sym} \left(g_{j,1} \otimes \cdots \otimes g_{j,k-1} \otimes g_{j,k+1} \otimes \cdots \otimes g_{j,n} \right) \right)_s \circ dW(g_{j,k})_s \\
&= \frac{1}{n} \sum_{j=1}^m \sum_{k=1}^n \int_0^t \frac{\prod_{l=1, l \neq k}^n W(g_{j,l})_s}{(n-1)!} \circ dW(g_{j,k})_s \\
&= \frac{1}{n!} \sum_{j=1}^m \prod_{k=1}^n W(g_{j,k})_t,
\end{aligned}
$$

which completes the proof. $\square$

7.3. **Proof of results in Section 4.**

7.3.1. *Proof of results in Section 4.1.*

Proof of Proposition 4.1. Let $p \in [1, \infty)$ and $q \in (1, \infty]$ with $\frac{1}{p} + \frac{1}{q} = 1$, where $\frac{1}{\infty} := 0$. Now, we assume by contradiction that $\mathrm{Pol}(X)$ defined in (9) is not dense in $L^p(\Omega, \mathcal{F}_T, \mathbb{P})$. Then, by the Hahn-Banach theorem, there exists a non-zero continuous linear functional $l : L^p(\Omega, \mathcal{F}_T, \mathbb{P}) \to \mathbb{R}$ that vanishes on $\mathrm{Pol}(X)$. Moreover, by the Riesz representation theorem, there exists some $Z \in L^q(\Omega, \mathcal{F}_T, \mathbb{P})$ such that for every $Y \in L^p(\Omega, \mathcal{F}_T, \mathbb{P})$ it holds that $l(Y) = \mathbb{E}[ZY]$. Hence, for every $Y \in \mathrm{Pol}(X)$, we have $l(Y) = \mathbb{E}[ZY] = 0$, e.g., $l(1) = \mathbb{E}[Z] = 0$. Thus, it suffices to show that $Z = 0 \in L^q(\Omega, \mathcal{F}_T, \mathbb{P})$, which contradicts the assumption that the continuous linear functional $l : L^p(\Omega, \mathcal{F}_T, \mathbb{P}) \to \mathbb{R}$ is non-zero.

To this end, we define for every fixed $m \in \mathbb{N}$, $i_{1:m} := (i_1, ..., i_m) \in \{1, ..., d\}^m$, and $t_{1:m} := (t_1, ..., t_m) \in [0, T]^m$ the random vector $\Omega \ni \omega \mapsto X_{t_{1:m}}^{i_{1:m}}(\omega) := \left(X_{t_1}^{i_1}(\omega), ..., X_{t_m}^{i_m}(\omega) \right)^\top \in \mathbb{R}^m$ and show that $\mathbb{E}\left[Z f\left(X_{t_{1:m}}^{i_{1:m}} \right) \right] = 0$ for all $f \in \mathcal{S}(\mathbb{R}^m; \mathbb{C})$, where $\mathcal{S}(\mathbb{R}^m; \mathbb{C})$ denotes the Schwartz space (see [22, p. 330]). For this purpose, we define the tempered distribution $\left(f \mapsto U(f) := \mathbb{E}\left[Z f\left(X_{t_{1:m}}^{i_{1:m}} \right) \right] \right) \in \mathcal{S}'(\mathbb{R}^m; \mathbb{C})$, where $\mathcal{S}'(\mathbb{R}^m; \mathbb{C})$ denotes the dual space of $\mathcal{S}(\mathbb{R}^m; \mathbb{C})$ (see [22, p. 332]). Moreover, we denote by $\left(f \mapsto \widehat{U}(f) := U(\widehat{f}) \right) \in \mathcal{S}'(\mathbb{R}^m; \mathbb{C})$ its Fourier transform in the sense of distributions (see [22, Equation 9.28]), where $\mathbb{R}^m \ni \lambda \mapsto \widehat{f}(\lambda) := \int_{\mathbb{R}^m} e^{-i\lambda^\top x} f(x) dx \in \mathbb{R}$ is the Fourier transform of $f \in \mathcal{S}(\mathbb{R}^m; \mathbb{C})$ (see [22, Equation 7.1]). Then, Fubini's theorem implies for every $f \in \mathcal{S}(\mathbb{R}^m; \mathbb{C})$ that

$$
\widehat{U}(f) = U(\widehat{f}) = \mathbb{E}\left[Z \widehat{f}\left(X_{t_{1:m}}^{i_{1:m}} \right) \right] = \mathbb{E}\left[Z \int_{\mathbb{R}^m} e^{-\mathbf{i} \left(X_{t_{1:m}}^{i_{1:m}} \right)^\top x} f(x) dx \right] = \int_{\mathbb{R}^m} \widehat{u}(x) f(x) dx, \tag{30}
$$

where $\mathbb{R}^m \ni x \mapsto \widehat{u}(x) := \mathbb{E}\left[Z e^{-\mathbf{i} x^\top X_{t_{1:m}}^{i_{1:m}}} \right] \in \mathbb{C}$. Moreover, by using the Cauchy-Schwarz inequality, that $\left\| X_{t_{1:m}}^{i_{1:m}} \right\| \leq \sum_{k=1}^m \left| X_{t_k}^{i_k} \right| \leq \sum_{k=1}^m \| X_{t_k} \|$, the generalized Hölder inequality (with exponents $\frac{1}{q} + \sum_{k=1}^m \frac{1}{mp} = 1$), and that X satisfies Assumption 2.1 (with $\varepsilon > 0$), we conclude for every $z \in V := \left\{ z \in \mathbb{C}^m : \| \mathrm{Im}(z) \| < \frac{\varepsilon}{mp} \right\}$ (with $\mathrm{Im}(z) := (\mathrm{Im}(z_1), ..., \mathrm{Im}(z_m))^\top$, for $z \in \mathbb{C}^m$) that

$$
\begin{aligned}
\mathbb{E}\left[\left| Z e^{-\mathbf{i} z^\top X_{t_{1:m}}^{i_{1:m}}} \right| \right] &\leq \mathbb{E}\left[|Z| e^{\| \mathrm{Im}(z) \| \left\| X_{t_{1:m}}^{i_{1:m}} \right\|} \right]^{\frac{1}{p}} \leq \mathbb{E}\left[|Z| e^{\| \mathrm{Im}(z) \| \sum_{k=1}^m \| X_{t_k} \|} \right]^{\frac{1}{p}} \\
&\leq \| Z \|_{L^q(\mathbb{P})} \prod_{k=1}^m \mathbb{E}\left[e^{mp \| \mathrm{Im}(z) \| \| X_{t_k} \|} \right]^{\frac{1}{mp}} \leq \| Z \|_{L^q(\mathbb{P})} \prod_{k=1}^m \mathbb{E}\left[e^{\varepsilon \| X_{t_k} \|} \right]^{\frac{1}{mp}} < \infty.
\end{aligned}
$$

Therefore, the function $V \ni z \mapsto \widehat{v}(z) := \mathbb{E}\left[Z e^{-\mathbf{i} z^\top X_{t_{1:m}}^{i_{1:m}}} \right] \in \mathbb{R}$ is holomorphic on V, which implies that the restriction $\widehat{u} := \widehat{v}|_{\mathbb{R}^m} : \mathbb{R}^m \to \mathbb{C}$ is real analytic. Moreover, by using that $\mathbb{E}[ZY] = 0$ for any

$Y \in \operatorname{Pol}(X)$, it follows for every $n \in \mathbb{N}_0$ and $(j_1, ..., j_n) \in \{1, ..., d\}^n$ that

$$\frac{\partial^n \widehat{u}}{\partial z_{j_1} \cdots \partial z_{j_n}}(0) = (-\mathbf{i})^n \mathbb{E}\left[Z X_{t_{j_1}}^{i_{j_1}} \cdots X_{t_{j_n}}^{i_{j_n}} \right] = 0.$$

Hence, by using that $\widehat{u} : \mathbb{R}^m \to \mathbb{C}$ is real analytic, the function $\widehat{u} : \mathbb{R}^m \to \mathbb{C}$ is constant equal to $\widehat{u}(0) = \mathbb{E}[Z] = 0$, which implies together with (30) that $\widehat{U}(f) = \int_{\mathbb{R}^m} \widehat{u}(x) f(x) dx = \widehat{u}(0) \int_{\mathbb{R}^m} f(x) dx = 0$ for all $f \in \mathcal{S}(\mathbb{R}^m; \mathbb{C})$. Thus, by using that the Fourier transform is bijective on $\mathcal{S}'(\mathbb{R}^m; \mathbb{C})$, it follows that $U = 0 \in \mathcal{S}'(\mathbb{R}^m; \mathbb{C})$, which shows that $\mathbb{E}\big[Z f\big(X_{t_{1:m}}^{i_{1:m}} \big) \big] = 0$ for all $f \in \mathcal{S}(\mathbb{R}^m; \mathbb{C})$.

Finally, we show that $Z = 0 \in L^q(\Omega, \mathcal{F}_T, \mathbb{P})$. To this end, we fix some $m \in \mathbb{N}$, $i_{1:m} := (i_1, ..., i_m) \in \{1, ..., d\}^m$, $t_{1:m} := (t_1, ..., t_m) \in [0,T]^m$, and $F \in \mathcal{B}(\mathbb{R}^m)$ and define again the random vector $\Omega \ni \omega \mapsto X_{t_{1:m}}^{i_{1:m}}(\omega) := \big(X_{t_1}^{i_1}(\omega), ..., X_{t_m}^{i_m}(\omega) \big)^\top \in \mathbb{R}^m$. Then, by using Lusin's theorem (see [57, Theorem 3.14]) and a mollification argument, the indicator function $\mathbb{1}_F : \mathbb{R}^m \to \mathbb{R}$ can be approximated by a sequence of smooth compactly supported functions $(f_n)_{n \in \mathbb{N}} \subseteq C_c^\infty(\mathbb{R}^m) \subseteq \mathcal{S}(\mathbb{R}^m; \mathbb{C})$ with respect to $\| \cdot \|_{L^p(\mathbb{P} \circ (X_{t_{1:m}}^{i_{1:m}})^{-1})}$, where $\mathcal{B}(\mathbb{R}^m) \ni \widetilde{F} \mapsto \big(\mathbb{P} \circ (X_{t_{1:m}}^{i_{1:m}})^{-1} \big) \big[\widetilde{F} \big] := \mathbb{P}\big[X_{t_{1:m}}^{i_{1:m}} \in \widetilde{F} \big] \in [0,1]$ denotes the push-forward measure. Hence, by using the triangle inequality, that $\mathbb{E}\big[Z f\big(X_{t_{1:m}}^{i_{1:m}} \big) \big] = 0$ for any $f \in \mathcal{S}(\mathbb{R}^m; \mathbb{C})$, and Hölder's inequality, we conclude that

$$\begin{aligned}
\left| \mathbb{E}\left[Z \mathbb{1}_F\big(X_{t_{1:m}}^{i_{1:m}} \big) \right] \right| &\leq \lim_{n \to \infty} \left| \mathbb{E}\left[Z (\mathbb{1}_F - f_n)\big(X_{t_{1:m}}^{i_{1:m}} \big) \right] \right| + \lim_{n \to \infty} \Big| \underbrace{\mathbb{E}\left[Z f_n\big(X_{t_{1:m}}^{i_{1:m}} \big) \right]}_{=0} \Big| \\
&\leq \|Z\|_{L^q(\mathbb{P})} \lim_{n \to \infty} \big\| (\mathbb{1}_F - f_n)\big(X_{t_{1:m}}^{i_{1:m}} \big) \big\|_{L^p(\mathbb{P})} \\
&= \|Z\|_{L^q(\mathbb{P})} \lim_{n \to \infty} \| \mathbb{1}_F - f_n \|_{L^p(\mathbb{P} \circ (X_{t_{1:m}}^{i_{1:m}})^{-1})} = 0.
\end{aligned} \tag{31}$$

Since $\mathbb{F} := (\mathcal{F}_t)_{t \in [0,T]}$ is the usual $\mathbb{P}$-augmented filtration[3] generated by X, the σ-algebra $\mathcal{F}_T$ is equal to the $\mathbb{P}$-completion of $\sigma(\{ X_t : t \in [0,T] \})$. Thus, by using that X has continuous sample paths, $\mathcal{F}_T$ is equal to the $\mathbb{P}$-completion of $\sigma\big(\bigcup_{m \in \mathbb{N}} \bigcup_{i_{1:m} \in \{1,...,d\}^m} \bigcup_{t_{1:m} \in [0,T]^m} \big\{ \big(X_{t_{1:m}}^{i_{1:m}} \big)^{-1}(F) : F \in \mathcal{B}(\mathbb{R}^m) \big\} \big)$. Hence, by applying the Dynkin π-λ theorem (see [18, Theorem 2.1.2], with corresponding π-system $\bigcup_{m \in \mathbb{N}} \bigcup_{i_{1:m} \in \{1,...,d\}^m} \bigcup_{t_{1:m} \in [0,T]^m} \big\{ \big(X_{t_{1:m}}^{i_{1:m}} \big)^{-1}(F) : F \in \mathcal{B}(\mathbb{R}^m) \big\}$ and λ-system $\sigma\big(\big\{ E \in \mathcal{F}_T : \mathbb{E}[Z \mathbb{1}_E] = 0 \big\} \big)$), we conclude from (31) that $\mathbb{E}[Z \mathbb{1}_E] = 0$ for all $E \in \mathcal{F}_T$. This shows that $Z = 0$, $\mathbb{P}$-a.s., and thus $Z = 0 \in L^q(\Omega, \mathcal{F}_T, \mathbb{P})$. □

Proof of Theorem 4.2. For (i), we fix some $g_n = \sum_{j=1}^{m_n} g_{n,j}^{\otimes n} \in L_{\text{diag}}^{np}(X)^{\otimes n}$. Then, by using the convention $g_{n,j}^{\otimes 0} := 1$ (if $n = 1$), Remark 3.8, and Proposition 3.10, it follows $\mathbb{P}$-a.s. that

$$\begin{aligned}
J_n^\circ(g_n)_T &= \sum_{j=1}^{m_n} J_n^\circ\left(g_{n,j}^{\otimes n} \right)_T = \sum_{j=1}^{m_n} \int_0^T J_{n-1}^\circ\left(g_{n,j}^{\otimes(n-1)} \right)_t \circ dW(g_{n,j})_t \\
&= \sum_{j=1}^{m_n} \int_0^T J_{n-1}^\circ\left(g_{n,j}^{\otimes(n-1)} \right)_t dW(g_{n,j})_t + \frac{1}{2} \sum_{j=1}^{m_n} \int_0^T J_{n-1}^\circ\left(g_{n,j}^{\otimes(n-1)} \right)_t g_{n,j}(t)^\top d\langle X \rangle_t g_{n,j}(t) \\
&= \sum_{j=1}^{m_n} \int_0^T \frac{W(g_{n,j})_t^{n-1}}{(n-1)!} g_{n,j}(t)^\top dX_t + \frac{1}{2} \sum_{j=1}^{m_n} \int_0^T \frac{W(g_{n,j})_t^{n-1}}{(n-1)!} g_{n,j}(t)^\top d\langle X \rangle_t g_{n,j}(t),
\end{aligned}$$

which completes the proof of (i).

For (ii), let $p \in [1,\infty)$ and $q \in (1,\infty]$ with $\frac{1}{p} + \frac{1}{q} = 1$, where $\frac{1}{\infty} := 0$. Now, we assume by contradiction that $\bigoplus_{n \in \mathbb{N}_0} \big\{ J_n^\circ(g_n)_T : g_n \in L_{\text{diag}}^{np}(X)^{\otimes n} \big\}$ is not dense in $L^p(\Omega, \mathcal{F}_T, \mathbb{P})$. Then, by the Hahn-Banach theorem, there exists a non-zero continuous linear functional $l : L^p(\Omega, \mathcal{F}_T, \mathbb{P}) \to \mathbb{R}$ that vanishes on $\bigoplus_{n \in \mathbb{N}_0} \big\{ J_n^\circ(g_n)_T : g_n \in L_{\text{diag}}^{np}(X)^{\otimes n} \big\}$. Moreover, by the Riesz representation theorem, there exists some $Z \in L^q(\Omega, \mathcal{F}_T, \mathbb{P})$ such that for every $Y \in L^p(\Omega, \mathcal{F}_T, \mathbb{P})$ it holds that $l(Y) = \mathbb{E}[ZY]$, e.g., $l(Y) = \mathbb{E}[ZY] = 0$ for all $Y \in \bigoplus_{n \in \mathbb{N}_0} \big\{ J_n^\circ(g_n)_T : g_n \in L_{\text{diag}}^{np}(X)^{\otimes n} \big\}$. Thus, it suffices to show that $Z = 0 \in L^q(\Omega, \mathcal{F}_T, \mathbb{P})$, which contradicts the assumption that $l : L^p(\Omega, \mathcal{F}_T, \mathbb{P}) \to \mathbb{R}$ is non-zero.

To this end, we fix some $m \in \mathbb{N}$, $(i_1, ..., i_m) \in \{1, ..., d\}^m$, $\mathbf{k} := (k_1, ..., k_m) \in \mathbb{N}_0^m$, and $(t_1, ..., t_m) \in [0,T]^m$. Moreover, we recall that X satisfies Assumption 2.1 (with $\varepsilon > 0$) and let $\delta > 0$ such that

$\delta \leq \frac{\varepsilon}{2mp}$. In addition, for every fixed $\lambda := (\lambda_1, ..., \lambda_m)^\top \in (-\delta, \delta)^m$, we define the simple function

$$[0,T] \ni t \quad \mapsto \quad g_\lambda(t) := \sum_{l=1}^m \lambda_l \mathbb{1}_{[0,t_l]}(t) e_{i_l} \in \mathbb{R}^d,$$

where e_i denotes the i-th unit vector of $\mathbb{R}^d$, $i = 1, ..., d$. Then, for every $\lambda := (\lambda_1, ..., \lambda_m)^\top \in (-\delta, \delta)^m$ and $t \in [0,T]$, we observe that

$$W(g_\lambda)_t = \sum_{l=1}^m \lambda_l \int_0^t \mathbb{1}_{[0,t_l]}(s) e_{i_l}^\top dX_s = \sum_{l=1}^m \lambda_l \left(X_{t \wedge t_l}^{i_l} - X_0^{i_l} \right). \tag{32}$$

Furthermore, by using that $g_\lambda : [0,T] \to \mathbb{R}^d$ is càglàd, Lemma 3.2 (ii) ensures that $g_\lambda \in L^{np}(X)$, for all $n \in \mathbb{N}$. Furthermore, by using the identity (32), that X_0 is $\mathbb{P}$-a.s. constant equal to some $x_0 \in \mathbb{R}^d$, Hölder's inequality, that $\delta mp \leq \varepsilon$, and that X satisfies Assumption 2.1 (with $\varepsilon > 0$), it holds that

$$\begin{aligned}
\mathbb{E}\left[|Z| e^{|W(g_\lambda)_T|} \right] &\leq \mathbb{E}\left[|Z| e^{\sum_{l=1}^m |\lambda_l| \left| X_{t_l}^{i_l} - X_0^{i_l} \right|} \right] \leq e^{\delta m \|x_0\|} \max_{l=1,\dots,m} \mathbb{E}\left[|Z| e^{\delta m \|X_{t_l}\|} \right] \\
&\leq \|Z\|_{L^q(\mathbb{P})} e^{\delta m \|x_0\|} \max_{l=1,\dots,m} \mathbb{E}\left[e^{\delta mp \|X_{t_l}\|} \right]^{\frac{1}{p}} \leq \|Z\|_{L^q(\mathbb{P})} e^{\delta m \|x_0\|} \max_{l=1,\dots,m} \mathbb{E}\left[e^{\varepsilon \|X_{t_l}\|} \right]^{\frac{1}{p}} < \infty.
\end{aligned} \tag{33}$$

Hence, by using that $\sum_{n=0}^N \frac{W(g_\lambda)_T^n}{n!}$ converges pointwise to $e^{W(g_\lambda)_T}$ together with $\left| Z \sum_{n=0}^N \frac{W(g_\lambda)_T^n}{n!} \right| \leq |Z| \sum_{n=0}^N \frac{|W(g_\lambda)_T|^n}{n!} \leq |Z| e^{|W(g_\lambda)_T|}$, where the latter is by (33) integrable, we can use the dominated convergence theorem, Proposition 3.10, and that $J_n^\circ \left(g_\lambda^{\otimes n} \right)_T \in \bigoplus_{n \in \mathbb{N}_0} \left\{ J_n^\circ (g_n)_T : g_n \in L^{np}_{\text{diag}}(X)^{\otimes n} \right\}$ to conclude that

$$u(\lambda) := \mathbb{E}\left[Z e^{W(g_\lambda)_T} \right] = \lim_{N \to \infty} \mathbb{E}\left[Z \sum_{n=0}^N \frac{W(g_\lambda)_T^n}{n!} \right] = \lim_{N \to \infty} \sum_{n=0}^N \mathbb{E}\left[Z J_n^\circ \left(g_\lambda^{\otimes n} \right)_T \right] = 0. \tag{34}$$

Since $\lambda \in (-\delta, \delta)^m$ was chosen arbitrarily, this shows that $u : (-\delta, \delta)^m \to \mathbb{R}$ vanishes on $(-\delta, \delta)^m$.

Next, we compute the $|\mathbf{k}|$-th order partial derivative of $u : (-\delta, \delta)^m \to \mathbb{R}$, k_l-times into the direction e_l, $l = 1, ..., m$, and aim to show that we can interchange them with the expectation of the random function $(-\delta, \delta)^m \times \Omega \ni (\lambda, \omega) \mapsto U(\lambda, \omega) := Z e^{W(g_\lambda)_T(\omega)} \in \mathbb{R}$, i.e. that

$$\partial_{\mathbf{k}} u(\lambda) := \frac{\partial^{|\mathbf{k}|} u}{\partial \lambda_1^{k_1} \cdots \partial \lambda_m^{k_m}}(0) = \mathbb{E}\left[\frac{\partial^{|\mathbf{k}|} U}{\partial \lambda_1^{k_1} \cdots \partial \lambda_m^{k_m}}(0, \cdot) \right] =: \mathbb{E}\left[\partial_{\mathbf{k}} U(0, \cdot) \right]. \tag{35}$$

If (35) holds true, then we can immediately conclude from (34) and (32) that

$$\begin{aligned}
0 &= \frac{\partial^{|\mathbf{k}|} u}{\partial \lambda_1^{k_1} \cdots \partial \lambda_m^{k_m}}(0) = \mathbb{E}\left[\frac{\partial^{|\mathbf{k}|} U}{\partial \lambda_1^{k_1} \cdots \partial \lambda_m^{k_m}}(0, \cdot) \right] = \mathbb{E}\left[\frac{\partial^{|\mathbf{k}|}}{\partial \lambda_1^{k_1} \cdots \partial \lambda_m^{k_m}} \left(Z e^{W(g_\lambda)_T} \right) \Big|_{\lambda=0} \right] \\
&= \mathbb{E}\left[Z \frac{\partial^{|\mathbf{k}|}}{\partial \lambda_1^{k_1} \cdots \partial \lambda_m^{k_m}} \left(\prod_{l=1}^m e^{\lambda_l \left(X_{t_l}^{i_l} - X_0^{i_l} \right)} \right) \Big|_{\lambda=0} \right] = \mathbb{E}\left[Z \prod_{l=1}^m \left(e^{\lambda_l \left(X_{t_l}^{i_l} - X_0^{i_l} \right)} \left(X_{t_l}^{i_l} - X_0^{i_l} \right)^{k_l} \right) \Big|_{\lambda=0} \right] \\
&= \mathbb{E}\left[Z \left(X_{t_1}^{i_1} - X_0^{i_1} \right)^{k_1} \cdots \left(X_{t_m}^{i_m} - X_0^{i_m} \right)^{k_m} \right].
\end{aligned} \tag{36}$$

In order to prove (35), we first use again (32), the inequality $e^{\delta|x|} |x|^j = \frac{j!}{\delta^j} e^{\delta|x|} \frac{|\delta x|^j}{j!} \leq \frac{j!}{\delta^j} e^{2\delta|x|}$ for any $j \in \mathbb{N}_0$ and $x \in \mathbb{R}$, that X_0 is $\mathbb{P}$-a.s. constant equal to $x_0 \in \mathbb{R}^d$, Hölder's inequality, that $2\delta mp \leq \varepsilon$

together with X satisfying Assumption 2.1 (with $\varepsilon > 0$) to obtain for every $\mathbf{j} := (j_1, ..., j_m) \in \mathbb{N}_0^m$ that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\widetilde{\lambda}\in(-\delta,\delta)^m}\left|\partial_{\mathbf{j}}U(\widetilde{\lambda},\cdot)\right|\right] = \mathbb{E}\left[\sup_{\widetilde{\lambda}\in(-\delta,\delta)^m}\left|\frac{\partial^{|\mathbf{j}|}}{\partial\lambda_1^{j_1}\cdots\partial\lambda_m^{j_m}}\left(Ze^{W(g_\lambda)_T}\right)\Big|_{\lambda=\widetilde{\lambda}}\right|\right] \\
&= \mathbb{E}\left[|Z|\sup_{\widetilde{\lambda}\in(-\delta,\delta)^m}\left|\prod_{l=1}^m\left(e^{\widetilde{\lambda}_l\left(X_{t_l}^{i_l}-X_0^{i_l}\right)}\left(X_{t_l}^{i_l}-X_0^{i_l}\right)^{j_l}\right)\right|\right] \leq \frac{\prod_{l=1}^m j_l!}{\delta^{|\mathbf{j}|}}\mathbb{E}\left[|Z|e^{2\delta\sum_{l=1}^m\left|X_{t_l}^{i_l}-X_0^{i_l}\right|}\right] \\
&\leq \frac{\prod_{l=1}^m j_l!}{\delta^{|\mathbf{j}|}}e^{2\delta m\|x_0\|}\max_{l=1,...,m}\mathbb{E}\left[|Z|e^{2\delta m\|X_{t_l}\|}\right] \leq \frac{\prod_{l=1}^m j_l!}{\delta^{|\mathbf{j}|}}\|Z\|_{L^q(\mathbb{P})}e^{2\delta m\|x_0\|}\max_{l=1,...,m}\mathbb{E}\left[e^{\varepsilon\|X_{t_l}\|}\right]^{\frac{1}{p}} < \infty.
\end{aligned} \tag{37}$$

Now, we show the identity (36) by induction on $|\mathbf{j}| = 0, ..., |\mathbf{k}|$, i.e. for every $\mathbf{j} := (j_1, ..., j_m) \in \mathbb{N}_0^m$ with $j_l \leq k_l$ for all $l = 1, ..., m$ we show that for every $\lambda \in (-\delta, \delta)^m$ it holds that

$$\partial_{\mathbf{j}}u(\lambda) = \mathbb{E}\left[\partial_{\mathbf{j}}U(\lambda,\cdot)\right]. \tag{38}$$

For the induction initialization $\mathbf{j} = (0, ..., 0)$ with $|\mathbf{j}| = 0$, there is nothing to prove. For the induction step, we assume without loss of generality that (38) holds true for some $\mathbf{j} := (j_1, ..., j_m) \in \mathbb{N}_0^m$ with $j_{l_0} < k_{l_0}$ for some $l_0 = 1, ..., m$, $j_l = 0$ for all $l = 1, ..., l_0 - 1$, and $j_l = k_l$ for all $l = l_0 + 1, ..., m$. Then, for every fixed $(\lambda_1, ..., \lambda_{l_0-1}, \lambda_{l_0+1}, ..., \lambda_m) \in \mathbb{R}^{m-1}$ and $\omega \in \Omega$, the functions $(-\delta, \delta) \ni \lambda_{l_0} \mapsto \frac{\partial}{\partial\lambda_{l_0}}\partial_{\mathbf{j}}U\big((\lambda_1, ..., \lambda_{l_0-1}, \lambda_{l_0}, \lambda_{l_0+1}, ..., \lambda_m), \omega\big)$ as well as $(-\delta, \delta) \ni \lambda_{l_0} \mapsto \mathbb{E}\big[\frac{\partial}{\partial\lambda_{l_0}}\partial_{\mathbf{j}}U\big((\lambda_1, ..., \lambda_{l_0-1}, \lambda_{l_0}, \lambda_{l_0+1}, ..., \lambda_m), \cdot\big)\big]$ exist and are continuous. Moreover, by using (37), we observe for every $\lambda_{l_0} \in (-\delta, \delta)$ that

$$\mathbb{E}\left[\left|\partial_{\mathbf{j}}U\big((\lambda_1, ..., \lambda_{l_0-1}, \lambda_{l_0}, \lambda_{l_0+1}, ..., \lambda_m), \cdot\big)\right|\right] \leq \mathbb{E}\left[\sup_{\widetilde{\lambda}\in(-\delta,\delta)^m}\left|\partial_{\mathbf{j}}U(\widetilde{\lambda},\cdot)\right|\right] < \infty.$$

as well as

$$\mathbb{E}\left[\sup_{\lambda_{l_0}\in(-\delta,\delta)}\left|\frac{\partial}{\partial\lambda_{l_0}}\partial_{\mathbf{j}}U\big((\lambda_1, ..., \lambda_{l_0-1}, \lambda_{l_0}, \lambda_{l_0+1}, ..., \lambda_m), \cdot\big)\right|\right] \leq \mathbb{E}\left[\sup_{\widetilde{\lambda}\in(-\delta,\delta)^m}\left|\partial_{\mathbf{j}+e_{l_0}}U(\widetilde{\lambda},\cdot)\right|\right] < \infty,$$

where $e_{l_0} := (0, ..., 0, 1, 0, ..., 0) \in \mathbb{N}_0^m$ with 1 at position l_0. Hence, we can apply [18, Theorem A.5.2] to conclude that (38) holds true with $\mathbf{j} + e_{l_0}$ instead of $\mathbf{j}$, proving the induction step and therefore (35).

Finally, we show that $Z = 0 \in L^q(\Omega, \mathcal{F}_T, \mathbb{P})$. To this end, we fix some $m \in \mathbb{N}$, $(i_1, ..., i_m) \in \{1, ..., d\}^m$, $(K_1, ..., K_m) \in \mathbb{N}_0^m$, and $(t_1, ..., t_m) \in [0, T]^m$. Then, by using m-times the binomial theorem (with $\binom{\mathbf{K}}{\mathbf{k}} := \frac{K_1!}{k_1!(K_1-k_1)!}\cdots\frac{K_m!}{k_m!(K_m-k_m)!}$), that X_0 is $\mathbb{P}$-a.s. constant equal to $x_0 \in \mathbb{R}^d$, and the identity (36), it follows that

$$\begin{aligned}
&\mathbb{E}\left[Z\left(X_{t_1}^{i_1}\right)^{K_1}\cdots\left(X_{t_m}^{i_m}\right)^{K_m}\right] = \mathbb{E}\left[Z\left(X_{t_1}^{i_1}-X_0^{i_1}+x_0^{i_1}\right)^{K_1}\cdots\left(X_{t_m}^{i_m}-X_0^{i_m}+x_0^{i_m}\right)^{K_m}\right] \\
&= \sum_{k_1=0}^{K_1}\cdots\sum_{k_m=0}^{K_m}\binom{\mathbf{K}}{\mathbf{k}}\mathbb{E}\left[Z\left(X_{t_1}^{i_1}-X_0^{i_1}\right)^{k_1}\cdots\left(X_{t_m}^{i_m}-X_0^{i_m}\right)^{k_m}\right]\left(x_0^{i_1}\right)^{K_1-k_1}\cdots\left(x_0^{i_m}\right)^{K_m-k_m} = 0.
\end{aligned}$$

However, since $\mathrm{Pol}(X)$ defined in (9) is by Proposition 4.1 dense in $L^p(\Omega, \mathcal{F}_T, \mathbb{P})$, we conclude from the Hahn-Banach theorem that $Z = 0 \in L^q(\Omega, \mathcal{F}_T, \mathbb{P})$. □

7.3.2. *Proof of results in Section 4.2.* For the results in this section, we denote by $\mathbb{G} := (\mathcal{G}_t)_{t\in[0,T]}$ the $\mathbb{P}$-completion of the filtration[6] generated by X. In particular, it holds for every $t \in [0, T)$ that $\bigcap_{s\in(t,T]}\mathcal{G}_s = \mathcal{F}_t$, whereas $\mathcal{G}_T = \mathcal{F}_T$. Moreover, we recall that the $\mathbb{F}$-predictable σ-algebra on $[0, T] \times \Omega$ coincides with the $\mathbb{G}$-predictable σ-algebra on $[0, T] \times \Omega$ (see [14, Theorem IV.97]). Furthermore, if for example the semimartingale X is a strong Markov process, then $\mathbb{G} = \mathbb{F}$ (see [35, Theorem 2.7.7]).

[6] The $\mathbb{P}$-completion of the filtration generated by X is defined as the smallest filtration $\mathbb{G} := (\mathcal{G}_t)_{t\in[0,T]}$ such that $\mathbb{G}$ is complete with respect to $(\Omega, \mathcal{A}, \mathbb{P})$ and X is $\mathbb{G}$-adapted.

Lemma 7.1. *Let X be a continuous semimartingale satisfying Assumption 2.1 and let $p \in [1,\infty)$. Then, the set of bounded elementary $\mathbb{F}$-predictable $\mathbb{R}^d$-valued processes*

$$\Theta_{e,d}^{\infty}(X) := \left\{ t \mapsto \sum_{k=1}^{K} \mathbb{1}_{(t_{k-1},t_k]}(t) Z_k : \begin{array}{c} K \in \mathbb{N},\, 0 = t_0 < t_1 < ... < t_K = T, \\ Z_k \in L^{\infty}(\Omega, \mathcal{G}_{t_{k-1}}, \mathbb{P}; \mathbb{R}^d),\, k = 1, ..., K \end{array} \right\} \tag{39}$$

is dense in $\Theta^p(X)$.

Proof. First, we observe that $\Theta_{e,d}^{\infty}(X) \subseteq \Theta^p(X)$ holds by definition. In order to see that $\Theta_{e,d}^{\infty}(X)$ is dense in $\Theta^p(X)$, we first show that any given $\theta := \left(\theta_t^1, ..., \theta_t^d\right)_{t\in[0,T]}^\top \in \Theta^p(X)$ can be approximated by the sequence of bounded $\mathbb{F}$-predictable processes $\theta^{(n)} := \left(\mathbb{1}_{\{|\theta_t^1| \leq n\}}\theta_t^1, ..., \mathbb{1}_{\{|\theta_t^d| \leq n\}}\theta_t^d\right)_{t\in[0,T]}^\top$, $n \in \mathbb{N}$, with respect to $\|\cdot\|_{\Theta^p(X)}$. Indeed, by applying the dominated convergence theorem twice, it follows that[7]

$$\begin{aligned}
&\lim_{n\to\infty} \left\|\theta - \theta^{(n)}\right\|_{\Theta^p(X)} = \lim_{n\to\infty} \sum_{i=1}^{d} \left\|\theta^i - \theta^{(i,n)}\right\|_{\Theta^p(X^i)} \\
&= \sum_{i=1}^{d} \lim_{n\to\infty} \mathbb{E}\left[\left(\int_0^T \mathbb{1}_{\{|\theta_t^i|>n\}} \left|\theta_t^i dA_t^i\right|\right)^p\right]^{\frac{1}{p}} + \sum_{i=1}^{d} \lim_{n\to\infty} \mathbb{E}\left[\left(\int_0^T \mathbb{1}_{\{|\theta_t^i|>n\}} \left(\theta_t^i\right)^2 d\langle M^i\rangle_t\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} \\
&= \sum_{i=1}^{d} \mathbb{E}\left[\left(\lim_{n\to\infty} \int_0^T \mathbb{1}_{\{|\theta_t^i|>n\}} \left|\theta_t^i dA_t^i\right|\right)^p\right]^{\frac{1}{p}} + \sum_{i=1}^{d} \mathbb{E}\left[\left(\lim_{n\to\infty} \int_0^T \mathbb{1}_{\{|\theta_t^i|>n\}} \left(\theta_t^i\right)^2 d\langle M^i\rangle_t\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} = 0,
\end{aligned}$$

where e_i denotes the i-th unit vector of $\mathbb{R}^d$. Hence, it suffices to show that every bounded $\mathbb{F}$-predictable process can be approximated by elementary $\mathbb{F}$-predictable processes from $\Theta_{e,d}^{\infty}(X)$.

Now, for every fixed $i = 1, ..., d$, we prove that the vector space of bounded $\mathbb{F}$-predictable real-valued processes is contained in $\mathcal{H}_i$ defined as the closure of $\Theta_{e,1}^{\infty}(X^i)$ (given as in (39) but with $d = 1$) with respect to $\|\cdot\|_{\Theta^p(X^i)}$. To this end, we first observe that $\mathcal{E}_T := \{\{0\} \times E : E \in \mathcal{G}_0\} \cup \{(s,t] \times E : 0 \leq s < t \leq T,\, E \in \mathcal{G}_s\}$ is a π-system with $[0,T] \times \Omega \in \mathcal{E}_T$, which generates the $\mathbb{F}$-predictable σ-algebra $\mathcal{P}_T := \sigma(\mathcal{E}_T)$ on $[0,T] \times \Omega$. Moreover, $\mathcal{H}_i$ is a vector space and for every $E \in \mathcal{E}_T$ it holds that $\mathbb{1}_E \in \mathcal{H}_i$. In addition, for every sequence $(\vartheta^{(i,n)})_{n\in\mathbb{N}} \subseteq \mathcal{H}_i$ with $0 \leq \vartheta_t^{(i,n)}(\omega) \leq \vartheta_t^{(i,n+1)}(\omega)$ for all $(t,\omega) \in [0,T] \times \Omega$ and $n \in \mathbb{N}$ converging pointwise to some bounded $\mathbb{F}$-predictable real-valued process $\vartheta^i := \left(\vartheta_t^i\right)_{t\in[0,T]}$, we use the dominated convergence theorem twice to obtain that

$$\begin{aligned}
&\lim_{n\to\infty} \left\|\vartheta_t^i - \vartheta_t^{(i,n)}\right\|_{\Theta^p(X^i)} \\
&\quad = \lim_{n\to\infty} \mathbb{E}\left[\left(\int_0^T \left|\left(\vartheta_t^i - \vartheta_t^{(i,n)}\right) dA_t^i\right|\right)^p\right]^{\frac{1}{p}} + \lim_{n\to\infty} \mathbb{E}\left[\left(\int_0^T \left(\vartheta_t^i - \vartheta_t^{(i,n)}\right)^2 d\langle M^i\rangle_t\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} \\
&\quad = \mathbb{E}\left[\left(\lim_{n\to\infty} \int_0^T \left|\left(\vartheta_t^i - \vartheta_t^{(i,n)}\right) dA_t^i\right|\right)^p\right]^{\frac{1}{p}} + \mathbb{E}\left[\left(\lim_{n\to\infty} \int_0^T \left(\vartheta_t^i - \vartheta_t^{(i,n)}\right)^2 d\langle M^i\rangle_t\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} = 0,
\end{aligned}$$

showing that $\vartheta^i \in \mathcal{H}_i$. Hence, we can apply the monotone class theorem (see [18, Theorem 6.1.3]) to conclude that $\mathcal{H}_i$ contains all bounded $\mathbb{F}$-predictable real-valued processes.

Finally, for every bounded $\mathbb{F}$-predictable process $\theta := \left(\theta_t^1, ..., \theta_t^d\right)_{t\in[0,T]}^\top$ and every $i = 1, ..., d$, there exists by the previous step a sequence $\left(\theta^{(i,n)}\right)_{n\in\mathbb{N}} \subseteq \Theta_{e,1}^{\infty}(X^i)$ with $\lim_{n\to\infty} \left\|\theta^i - \theta^{(i,n)}\right\|_{\Theta^p(X^i)} = 0$. Hence, by defining $\theta^{(n)} := \left(\theta_t^{(1,n)}, ..., \theta_t^{(d,n)}\right)_{t\in[0,T]}^\top \in \Theta_{e,d}^{\infty}(X)$, $n \in \mathbb{N}$, it follows that

$$\lim_{n\to\infty} \left\|\theta - \theta^{(n)}\right\|_{\Theta^p(X)} = \sum_{i=1}^{d} \lim_{n\to\infty} \left\|\theta^i - \theta^{(n,i)}\right\|_{\Theta^p(X^i)} = 0,$$

which completes the proof. $\square$

[7] For every $i = 1, ..., d$ the vector space $\Theta^p(X^i)$ consists of $\mathbb{F}$-predictable real-valued processes $\theta := (\theta_t)_{t\in[0,T]}$ such that $\|\theta\|_{\Theta^p(X^i)} := \mathbb{E}\left[\left(\int_0^T |\theta_t^i dA_t^i|\right)^p\right]^{1/p} + \mathbb{E}\left[\left(\int_0^T (\theta_t^i)^2 d\langle M^i\rangle_t\right)^{p/2}\right]^{1/p} < \infty$.

Corollary 7.2. *Let X be a continuous semimartingale satisfying Assumption 2.1, let $p \in [1,\infty)$, and $t \in [0,T]$. Then, the following holds true:*

(i) *The set of polynomials*

$$\mathrm{Pol}(X|_{[0,t]}) := \operatorname{span}\left\{ \left(X_{t_1}^{i_1}\right)^{k_1} \cdots \left(X_{t_m}^{i_m}\right)^{k_m} : \begin{array}{c} m \in \mathbb{N},\ (i_1,...,i_m) \in \{1,...,d\}^m, \\ (k_1,...,k_m) \in \mathbb{N}_0^m,\ (t_1,...,t_m) \in [0,t]^m \end{array} \right\}$$

is dense in $L^p(\Omega,\mathcal{G}_t,\mathbb{P})$.

(ii) *The direct sum $\bigoplus_{n\in\mathbb{N}_0} \left\{ J_n^\circ(g_n)_t : g_n \in L^{np}_{\mathrm{diag}}(X|_{[0,t]})^{\otimes n} \right\}$ is dense in $L^p(\Omega,\mathcal{G}_t,\mathbb{P})$.*

Proof. For (i), we can follow the proof of Proposition 4.1 (but with t replacing T) and use that $\mathcal{G}_t$ is defined as the $\mathbb{P}$-completion of $\sigma(\{X_s : s \in [0,t]\})$. For (ii), we can follow the proof of Theorem 4.2 (but with t replacing T), where we now apply (i) instead of Theorem 4.1. $\square$

Proposition 7.3. *Let X be a continuous semimartingale satisfying Assumption 2.1 and let $p \in [1,\infty)$. Then, $\mathcal{G}^p(X) := \left\{ \int_0^T \theta_t^\top dX_t : \theta \in \Theta^p(X) \right\}$ is contained in the closure of*

$$\mathcal{J}^p(X) := \left\{ \int_0^T \left(\vartheta_t^{g_{1:N}}\right)^\top dX_t : N, m_n \in \mathbb{N} \text{ and } g_{n,j,0}, g_{n,j,1} \in L^{np}(X),\ n = 1,...,N,\ j = 1,...,m_n \right\}$$

with respect to $\|\cdot\|_{L^p(\mathbb{P})}$, where $\vartheta_t^{g_{1:N}} := \sum_{n=1}^N \sum_{j=1}^{m_n} \frac{W(g_{n,j,0})_t^{n-1}}{(n-1)!} g_{n,j,1}(t)$, for $t \in [0,T]$.

Proof. Fix some $\theta \in \Theta^p(X)$ and $\varepsilon > 0$. Then, by using Lemma 7.1, there exists an elementary $\mathbb{F}$-predictable process $\widetilde{\theta} \in \Theta^\infty_{e,d}(X)$ of the form $t \mapsto \sum_{k=1}^K \mathbb{1}_{(t_{k-1},t_k]}(t) Z_k$, for some $K \in \mathbb{N}$, $0 = t_0 < t_1 < ... < t_K = T$, and $Z_k := (Z_{k,1},...,Z_{k,d})^\top \in L^\infty(\Omega,\mathcal{G}_{t_{k-1}},\mathbb{P};\mathbb{R}^d)$, $k = 1,...,K$, such that

$$\left\|\theta - \widetilde{\theta}\right\|_{\Theta^p(X)} < \frac{\varepsilon}{2\max(1,C_p)}, \tag{40}$$

where $C_p > 0$ is the constant in the upper Burkholder-Davis-Gundy inequality. Moreover, for every $k = 1,...,K$ and $i = 1,...,d$, we apply Corollary 7.2 (ii) on $Z_{k,i} \in L^\infty(\Omega,\mathcal{G}_{t_{k-1}},\mathbb{P}) \subseteq L^{2p}(\Omega,\mathcal{G}_{t_{k-1}},\mathbb{P})$ to obtain some $N, m_n \in \mathbb{N}$ and $h_{k,i,n,j} \in L^{2np}_{\mathrm{diag}}(X)$, $n = 0,...,N-1$ and $j = 1,...,m_n$, such that

$$\mathbb{E}\left[\left| Z_{k,i} - \sum_{n=0}^{N-1}\sum_{j=1}^{m_n} J_n^\circ\left(h_{k,i,n,j}^{\otimes n}\right)_{t_{k-1}} \right|^{2p}\right]^{\frac{1}{2p}} < \frac{\varepsilon}{2\max(1,C_p)C_{A,M,p}dK}, \tag{41}$$

where $C_{A,M,p} := 1 + \mathbb{E}\left[\left(\int_0^T \|dA_t\|\right)^{2p}\right]^{1/(2p)} + \mathbb{E}\left[\|\langle M\rangle_T\|_F^{p}\right]^{1/(2p)} > 0$. From this, we define for every $(k,i,j) \in \{1,...,K\} \times \{1,...,d\} \times \{1,...,m_n\}$ the functions $g_{1,(k,i,j),0}, g_{1,(k,i,j),1} \in L^{np}(X)$ by

$$\begin{aligned} [0,T] \ni t &\mapsto g_{1,(k,i,j),0}(t) := 0 \in \mathbb{R}^d \\ [0,T] \ni t &\mapsto g_{1,(k,i,j),1}(t) := h_{k,i,0,j}\mathbb{1}_{(t_{k-1},t_k]} e_i \in \mathbb{R}^d, \end{aligned}$$

where e_i denotes the i-th unit vector of $\mathbb{R}^d$. Moreover, for every $n = 2,...,N$ and $(k,i,j) \in \{1,...,K\} \times \{1,...,d\} \times \{1,...,m_n\}$, we define the functions $g_{n,(k,i,j),0}, g_{n,(k,i,j),1} \in L^{np}(X)$ by

$$\begin{aligned} [0,T] \ni t &\mapsto g_{n,(k,i,j),0}(t) := h_{k,i,n-1,j} \in \mathbb{R}^d \\ [0,T] \ni t &\mapsto g_{n,(k,i,j),1}(t) := \mathbb{1}_{(t_{k-1},t_k]} e_i \in \mathbb{R}^d \end{aligned}$$

In addition, for every $k = 1, ..., K$ and $i = 1, ..., d$, we define $\widetilde{Z}_{k,i} := \sum_{n=0}^{N-1} \sum_{j=1}^{m_n} J_n\big(h_{k,i,n,j}^{\otimes n}\big)_{t_{k-1}}$. Furthermore, we define $\vartheta^{g_{1:N}} := \big(\vartheta_t^{g_{1:N},1}, ..., \vartheta_t^{g_{1:N},1}\big)_{t\in[0,T]}^\top$ for every $i = 1, ..., d$ and $t \in [0, T]$ by

$$\begin{aligned}
\vartheta_t^{g_{1:N},i} &:= \sum_{n=1}^{N} \sum_{j=1}^{m_n} \frac{W\left(g_{n,(k,i,j),0}\right)_{t_{k-1}}^{n-1}}{(n-1)!} g_{n,(k,i,j),1}(t) \\
&= \sum_{j=1}^{m_n} h_{k,i,0,j} \mathbb{1}_{(t_{k-1},t_k]}(t) e_i + \sum_{n=2}^{N} \sum_{j=1}^{m_n} \frac{W\left(h_{k,i,n-1,j}\right)_{t_{k-1}}^{n-1}}{(n-1)!} \mathbb{1}_{(t_{k-1},t_k]}(t) e_i \\
&= \sum_{j=1}^{m_n} h_{k,i,0,j} \mathbb{1}_{(t_{k-1},t_k]}(t) e_i + \sum_{n=2}^{N} \sum_{j=1}^{m_n} J_{n-1}^\circ\left(h_{k,i,n-1,j}^{\otimes(n-1)}\right)_{t_{k-1}} \mathbb{1}_{(t_{k-1},t_k]}(t) e_i \\
&= \left(\sum_{n=0}^{N-1} \sum_{j=1}^{m_n} J_n^\circ\left(h_{k,i,n,j}^{\otimes n}\right)_{t_{k-1}}\right) \mathbb{1}_{(t_{k-1},t_k]}(t) e_i = \mathbb{1}_{(t_{k-1},t_k]}(t) \widetilde{Z}_{k,i} e_i,
\end{aligned} \tag{42}$$

where the third equality follows from Proposition 3.10. Hence, by using (42), that $\left|x^\top y\right| \leq \|x\|\|y\|$ and $x^\top \widetilde{M} x \leq \|x\|^2 \|\widetilde{M}\|_F$, Hölder's inequality, Minkowski's inequality together with $\big(\sum_{k=1}^K c_{k,i}\big)^{1/2} \leq \sum_{k=1}^K c_{k,i}^{1/2}$ for all $c_{k,i} \geq 0$, Proposition 3.10, and the inequality (41), it follows that

$$\begin{aligned}
&\left\|\widetilde{\theta} - \vartheta^{g_{1:N}}\right\|_{\Theta^p(X)} \\
&= \sum_{i=1}^{d} \mathbb{E}\left[\left(\int_0^T \left|\left(\widetilde{\theta}_t^i - \vartheta_t^{g_{1:N},i}\right) dA_t^i\right|\right)^p\right]^{\frac{1}{p}} + \sum_{i=1}^{d} \mathbb{E}\left[\left(\int_0^T \left(\widetilde{\theta}_t^i - \vartheta_t^{g_{1:N},i}\right)^2 d\langle M^i\rangle_t\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} \\
&= \sum_{i=1}^{d} \mathbb{E}\left[\left(\int_0^T \left|\left(\sum_{k=1}^{K} \mathbb{1}_{(t_{k-1},t_k]}(t) \left(Z_{k,i} - \widetilde{Z}_{k,i}\right)\right) dA_t^i\right|\right)^p\right]^{\frac{1}{p}} \\
&\quad + \sum_{i=1}^{d} \mathbb{E}\left[\left(\int_0^T \left(\sum_{k=1}^{K} \mathbb{1}_{(t_{k-1},t_k]}(t) \left(Z_{k,i} - \widetilde{Z}_{k,i}\right)\right)^2 d\langle M^i\rangle_t\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} \\
&\leq \sum_{i=1}^{d} \mathbb{E}\left[\left(\sum_{k=1}^{K} \left|Z_{k,i} - \widetilde{Z}_{k,i}\right| \int_{t_{k-1}}^{t_k} \left|dA_t^i\right|\right)^p\right]^{\frac{1}{p}} + \sum_{i=1}^{d} \mathbb{E}\left[\left(\sum_{k=1}^{K} \left|Z_{k,i} - \widetilde{Z}_{k,i}\right|^2 \left(\langle M^i\rangle_{t_k} - \langle M^i\rangle_{t_{k-1}}\right)\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} \\
&\leq \sum_{i=1}^{d} \sum_{k=1}^{K} \mathbb{E}\left[|Z_{k,i} - \widetilde{Z}_{k,i}|^{2p}\right] \mathbb{E}\left[\left(\int_0^T \|dA_t\|\right)^{2p}\right]^{\frac{1}{2p}} + \sum_{i=1}^{d} \sum_{k=1}^{K} \mathbb{E}\left[|Z_{k,i} - \widetilde{Z}_{k,i}|^{2p}\right] \mathbb{E}\left[\|\langle M\rangle_T\|_F^{\frac{p}{2}}\right]^{\frac{1}{2p}} \\
&\leq C_{A,M,p} \sum_{i=1}^{d} \sum_{k=1}^{K} \mathbb{E}\left[\left|Z_{k,i} - \sum_{n=0}^{N-1} \sum_{j=1}^{m_n} J_n^\circ\left(h_{k,i,n,j}^{\otimes n}\right)_{t_{k-1}}\right|^{2p}\right]^{\frac{1}{2p}} \\
&< C_{A,M,p} dK \frac{\varepsilon}{2\max(1, C_p) C_{A,M,p} dK} = \frac{\varepsilon}{2\max(1, C_p)}.
\end{aligned} \tag{43}$$

Thus, by using Minkowski's inequality, the Burkholder-Davis-Gundy inequality with constant $C_p > 0$, and the inequalities (40)+(43), it follows that

$$\begin{aligned}
\mathbb{E}\left[\left|\int_0^T \theta_t^\top dX_t - \int_0^T \left(\vartheta_t^{g_{1:N}}\right)^\top dX_t\right|^p\right]^{\frac{1}{p}} &\leq \sum_{i=1}^d \mathbb{E}\left[\left|\int_0^T \left(\theta_t^i - \vartheta_t^{g_{1:N},i}\right) dX_t^i\right|^p\right]^{\frac{1}{p}} \\
&\leq \sum_{i=1}^d \mathbb{E}\left[\left|\int_0^T \left(\theta_t^i - \vartheta_t^{g_{1:N},i}\right) dA_t^i\right|^p\right]^{\frac{1}{p}} + \sum_{i=1}^d \mathbb{E}\left[\left|\int_0^T \left(\theta_t^i - \vartheta_t^{g_{1:N},i}\right) dM_t^i\right|^p\right]^{\frac{1}{p}} \\
&\leq \sum_{i=1}^d \mathbb{E}\left[\left|\int_0^T \left(\theta_t^i - \vartheta_t^{g_{1:N},i}\right) dA_t^i\right|^p\right]^{\frac{1}{p}} + C_p \sum_{i=1}^d \mathbb{E}\left[\left(\int_0^T \left(\theta_t^i - \vartheta_t^{g_{1:N},i}\right)^2 d\langle M^i\rangle_t\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} \\
&\leq \max(1,C_p) \sum_{i=1}^d \left\|\theta^i - \vartheta^{g_{1:N},i}\right\|_{\Theta^p(X^i)} = \max(1,C_p)\left\|\theta - \vartheta^{g_{1:N}}\right\|_{\Theta^p(X)} \\
&\leq \max(1,C_p)\left\|\theta - \widetilde{\theta}\right\|_{\Theta^p(X)} + \max(1,C_p)\left\|\widetilde{\theta} - \vartheta^{g_{1:N}}\right\|_{\Theta^p(X)} \\
&< \max(1,C_p)\frac{\varepsilon}{2\max(1,C_p)} + \max(1,C_p)\frac{\varepsilon}{2\max(1,C_p)} = \varepsilon.
\end{aligned}$$

Since $\theta \in \Theta^p(X)$ and $\varepsilon > 0$ were chosen arbitrarily, this shows that $\mathcal{G}^p(X)$ is contained in the closure of $\mathcal{J}^p(X)$ with respect to $\|\cdot\|_{L^p(\mathbb{P})}$. $\square$

Proof of Theorem 4.4. Fix some $G \in L^p(\Omega, \mathcal{F}_T, \mathbb{P})$ and $\varepsilon > 0$. Then, by definition of the infimum, there exists some $(\widetilde{c}, \widetilde{\theta}) \in \mathbb{R} \times \Theta^p(X)$ such that

$$\left\|G - \widetilde{c} - \int_0^T \widetilde{\theta}_t^\top dX_t\right\|_{L^p(\mathbb{P})} - \frac{\varepsilon}{2} < + \inf_{(c,\theta)\in\mathbb{R}\times\Theta^p(X)} \left\|G - c - \int_0^T \theta_t^\top dX_t\right\|_{L^p(\mathbb{P})}. \tag{44}$$

Moreover, by applying Proposition 7.3 to $\int_0^T \widetilde{\theta}_t^\top dX_t \in \mathcal{G}^p(X)$, there exists some $N, m_n \in \mathbb{N}$ and $g_{n,j,0}, g_{n,j,1} \in L^{np}(X)$, $n = 1, ..., N$ and $j = 1, ..., m_n$ such that $\int_0^T \left(\vartheta_t^{g_{1:N}}\right)^\top dX_t$ with $\vartheta_t^{g_{1:N}} := \sum_{n=1}^N \sum_{j=1}^{m_n} \frac{W(g_{n,j,0})_t^{n-1}}{(n-1)!} g_{n,j,1}(t)$, $t \in [0,T]$, satisfies

$$\left\|\int_0^T \widetilde{\theta}_t^\top dX_t - \int_0^T \left(\vartheta_t^{g_{1:N}}\right)^\top dX_t\right\|_{L^p(\mathbb{P})} < \frac{\varepsilon}{2}. \tag{45}$$

Moreover, we set $g_0 := \widetilde{c} \in \mathbb{R}$. Hence, by using the inequality (44), the reverse triangle inequality of $\|\cdot\|_{L^p(\mathbb{P})}$, and the inequality (45), it follows that

$$\begin{aligned}
&\left\|G - g_0 - \int_0^T \left(\vartheta_t^{g_{1:N}}\right)^\top dX_t\right\|_{L^p(\mathbb{P})} - \inf_{(c,\theta)\in\mathbb{R}\times\Theta^p(X)} \left\|G - c - \int_0^T \theta_t^\top dX_t\right\|_{L^p(\mathbb{P})} \\
&\quad < \left\|G - g_0 - \int_0^T \left(\vartheta_t^{g_{1:N}}\right)^\top dX_t\right\|_{L^p(\mathbb{P})} - \left\|G - \widetilde{c} - \int_0^T \widetilde{\theta}_t^\top dX_t\right\|_{L^p(\mathbb{P})} + \frac{\varepsilon}{2} \\
&\quad \leq \left\|\int_0^T \widetilde{\theta}_t^\top dX_t - \int_0^T \left(\vartheta_t^{g_{1:N}}\right)^\top dX_t\right\|_{L^p(\mathbb{P})} + \frac{\varepsilon}{2} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,
\end{aligned}$$

which completes the proof. $\square$

7.4. **Proof of results in Section 5.**

7.4.1. *Proof of results in Section 5.1.*

Lemma 7.4. *Let X be a continuous semimartingale satisfying Assumption 2.1 and let $p \in [1,\infty)$. Then, $C([0,T];\mathbb{R}^d)$ is dense in $L^p(X)$.*

Proof. First, we show that the set of step functions is dense in $L^p(X)$, where a step function is of the form $[0,T] \ni t \mapsto \mathbb{1}_{\{0\}}(t)v_0 + \sum_{l=1}^m \mathbb{1}_{(t_{l-1},t_l]}(t)v_l \in \mathbb{R}^d$, for some $m \in \mathbb{N}$, $0 = t_0 < t_1 < ... < t_m = T$, and $v_0, ..., v_m \in \mathbb{R}^d$. To this end, we fix some $g \in L^p(X)$ and $\varepsilon > 0$. Then, by using that $g \in D_l([0,T];\mathbb{R}^d)$, i.e. that $[0,T] \ni t \mapsto g(T-t) \in \mathbb{R}^d$ is càdlàg, we can apply [1, Lemma 12.1] to conclude that there exists some $0 = t_0 < t_1 < ... < t_m = T$ and $v_1, ..., v_m \in \mathbb{R}^d$ such that for every $l = 1, ..., m$ it holds that $\sup_{u,v\in(t_{l-1},t_l]} \|g(u) - g(v)\| < C_{p,T,X}^{-1}\varepsilon$, where the constant $C_{p,T,X} > 0$ was introduced in Lemma 3.2 (ii). Hence, by using Lemma 3.2 (ii), it follows for the step function $[0,T] \ni t \mapsto s(t) := \mathbb{1}_{\{0\}}(t)g(0) + \sum_{l=1}^m \mathbb{1}_{(t_{l-1},t_l]}(t)g(t_l) \in \mathbb{R}^d$ that

$$\begin{aligned}\|g - s\|_{L^p(X)} &\leq C_{p,T,X}\|g - s\|_\infty \leq C_{p,T,X} \left(\|g(0) - s(0)\| + \max_{l=1,...,m} \sup_{u\in(t_{l-1},t_l]} \|g(u) - s(u)\| \right) \\ &= C_{p,T,X} \max_{l=1,...,m} \sup_{u\in(t_{l-1},t_l]} \|g(u) - g(t_l)\| < C_{p,T,X} \frac{\varepsilon}{C_{p,T,X}} = \varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ and $g \in L^p(X)$ were chosen arbitrarily, this shows that step functions are dense in $L^p(X)$.

Next, we show that every fixed step function $[0,T] \ni t \mapsto s(t) := \mathbb{1}_{\{0\}}(t)v_0 + \sum_{l=1}^m \mathbb{1}_{(t_{l-1},t_l]}(t)v_l \in \mathbb{R}^d$ can be approximated by some element in $C([0,T];\mathbb{R}^d)$ with respect to $\|\cdot\|_{L^p(X)}$, where $m \in \mathbb{N}$, $0 = t_0 < t_1 < ... < t_m = T$, and $v_0, ..., v_m \in \mathbb{R}^d$. To this end, we define for every fixed $\delta \in (0, t_{\min})$ with $t_{\min} := \min_{l=1,...,m}(t_l - t_{l-1})$ the function

$$[0,T] \ni t \quad \mapsto \quad f_\delta(t) := \begin{cases} v_{l-1}(1-\delta) + v_l\delta, & t \in (t_{l-1}, t_{l-1}+\delta], \\ v_l, & t \in (t_{l-1}+\delta, t_l]. \end{cases}$$

Then, $f_\delta : [0,T] \to \mathbb{R}^d$ is continuous with $\|f_\delta\|_\infty = \|s\|_\infty$ and converges pointwise to the fixed step function $s : [0,T] \to \mathbb{R}^d$, as $\delta \to 0$. Moreover, by using that X is a continuous semimartingale, there exists some $E \in \mathcal{A}$ with $\mathbb{P}[E] = 1$ such that for every $\omega \in E$ the paths $t \mapsto X_t(\omega)$, $t \mapsto A_t(\omega)$, and $t \mapsto \langle M\rangle_t(\omega)$ are continuous. Now, for every fixed $\omega \in E$, we use the dominated convergence theorem (with $f_\delta \to s$ pointwise, as $\delta \to 0$, and $\|f_\delta\|_\infty = \|s\|_\infty$ for all $\delta > 0$) to conclude that

$$\lim_{\delta\to 0} \int_0^T \|s(t) - f_\delta(t)\|\|dA_t(\omega)\| = 0 \quad \text{and} \quad \lim_{\delta\to 0} \int_0^T (s(t) - f_\delta(t))^\top d\langle M\rangle_t(\omega)(s(t) - f_\delta(t)) = 0. \tag{46}$$

In addition, by using (25), we observe for every $\delta \in (0,1)$ that, $\mathbb{P}$-a.s.,

$$\begin{aligned}\left(\int_0^T \|s(t) - f_\delta(t)\|\|dA_t\|\right)^p &\leq (2\|s\|_\infty)^p \left(\int_0^T \|dA_t\|\right)^p, \quad \text{and} \\ \left(\int_0^T (s(t) - f_\delta(t))^\top d\langle M\rangle_t(s(t) - f_\delta(t))\right)^{\frac{p}{2}} &\leq (3d)^{\frac{p}{2}}(2\|s\|_\infty)^p \|\langle M\rangle_T\|_F^{\frac{p}{2}},\end{aligned} \tag{47}$$

where both right-hand sides do not depend on $\delta \in (0,1)$ and are $\mathbb{P}$-integrable due to (24) (using that X satisfies Assumption 2.1). Hence, by another application of the dominated convergence theorem (with $\left(\int_0^T \|s(t) - f_\delta\|\|dA_t\|\right)^p \to 0$ and $\left(\int_0^T (s(t) - f_\delta(t))^\top d\langle M\rangle_t(s(t) - f_\delta(t))\right)^{p/2} \to 0$, as $\delta \to 0$, $\mathbb{P}$-a.s. (see (46)), and (47)), we have

$$\begin{aligned}\lim_{\delta\to 0} \|s - f_\delta\|_{L^p(X)} &\leq \lim_{\delta\to 0} \mathbb{E}\left[\left(\int_0^T \|s(t) - f_\delta(t)\|\|dA_t\|\right)^p\right]^{\frac{1}{p}} \\ &\quad + \lim_{\delta\to 0} \mathbb{E}\left[\left(\int_0^T (s(t) - f_\delta(t))^\top d\langle M\rangle_t(s(t) - f_\delta(t))\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} = 0,\end{aligned}$$

which completes the proof. □

Lemma 7.5. *Let X be a continuous semimartingale satisfying Assumption 2.1, let $p \in [1,\infty)$, and $\rho \in C(\mathbb{R})$ be non-polynomial. Then, $\mathcal{NN}_{d,1}^\rho$ is dense in $L^p(X)$.*

Proof. For $p \in [1,\infty)$, let $g \in L^p(X)$ and fix some $\varepsilon > 0$. Then, by using Lemma 7.4, there exists some $f := (f_1, ..., f_d)^\top \in C([0,T];\mathbb{R}^d)$ such that

$$\|g - f\|_{L^p(X)} < \frac{\varepsilon}{2}. \tag{48}$$

Moreover, for every fixed $i = 1, ..., d$, we apply the universal approximation theorem in [41, Theorem 1] to conclude that there exists a neural network $\mathbb{R} \ni t \mapsto \varphi_i(t) := \sum_{n=1}^N y_{n,i}\rho(a_{n,i}t + b_{n,i}) \in \mathbb{R}$, for some

$a_{n,i},, b_{n,i}, y_{n,i} \in \mathbb{R}$, $n = 1, ..., N$, such that

$$\|f_i - \varphi_i\|_\infty := \sup_{t\in[0,T]} |g_i(t) - \varphi_i(t)| < \frac{\varepsilon}{2C_{p,T,X}\sqrt{d}}, \tag{49}$$

where the constant $C_{p,T,X} > 0$ was introduced in Lemma 3.2 (ii). From this, we define the neural network $[0,T] \ni t \mapsto \varphi := (\varphi_1, ..., \varphi_d)^\top = \sum_{n=1}^N y_n \rho(a_n t + b_n) \in \mathbb{R}^d$ satisfying $\varphi \in \mathcal{NN}_{d,1}^\rho$, where $\rho \in C(\mathbb{R})$ is applied componentwise, and where $a_n := (a_{n,1}, ..., a_{n,d})^\top \in \mathbb{R}^d$, $b_n := (b_{n,1}, ..., b_{n,d})^\top \in \mathbb{R}^d$, and $y_n := (y_{n,1}, ..., y_{n,d})^\top \in \mathbb{R}^d$, $n = 1, ..., N$. Then, by using Lemma 3.2 (ii), Minkowski's integral inequality, and the inequalities (48)+(49), it follows that

$$\begin{aligned}
\|g - \varphi\|_{L^p(X)} &\leq \|g - f\|_{L^p(X)} + \|f - \varphi\|_{L^p(X)} \\
&\leq \|g - f\|_{L^p(X)} + C_{p,T,X} \|f - \varphi\|_\infty \\
&\leq \|g - f\|_{L^p(X)} + C_{p,T,X} \sup_{t\in[0,T]} \left(\sum_{i=1}^d |f_i(t) - \varphi_i(t)|^2 \right)^{\frac{1}{2}} \\
&\leq \|g - f\|_{L^p(X)} + C_{p,T,X} \sqrt{d} \max_{i=1,...,d} \sup_{t\in[0,T]} |f_i(t) - \varphi_i(t)| \\
&< \frac{\varepsilon}{2} + C_{p,T,X}\sqrt{d} \frac{\varepsilon}{2C_{p,T,X}\sqrt{d}} = \varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ and $g \in L^p(X)$ were chosen arbitrarily, this shows that $\mathcal{NN}_{d,1}^\rho$ is dense in $L^p(X)$. $\square$

Proof of Proposition 5.3. For $n = 0$ and $p \in [1, \infty)$, we first observe that $\mathcal{NN}_{d,0}^\rho := \mathbb{R}$ is dense in $L^{np}(X)^{\otimes 0} := \mathbb{R}$. Moreover, for $n = 1$, the conclusion follows from Lemma 7.5 by using the identification $L^{np}(X)^{\otimes 1} \cong L^p(X)$. Now, for $n \geq 2$, we fix some $\varepsilon \in (0,1)$ and $g \in L^{np}_{\text{diag}}(X)^{\otimes n}$, which has a representation $g = \sum_{j=1}^m g_j^{\otimes n}$, for some $g_1, ..., g_m \in L^{np}(X)$, such that

$$\left\| g - \sum_{j=1}^m g_j^{\otimes n} \right\|_{L^{np}(X)^{\otimes n}} < \frac{\varepsilon}{2}. \tag{50}$$

Then, for every fixed $j = 1, ..., m$, we apply Lemma 7.5 to obtain some $\varphi_j \in \mathcal{NN}_{d,1}^\rho$ such that

$$\|g_j - \varphi_j\|_{L^{np}(X)} < \frac{\varepsilon}{2mn \left(1 + \|g_j\|_{L^{np}(X)}\right)^{n-1}} \leq 1,$$

which implies that $\|\varphi_j\|_{L^{np}(X)} \leq 1 + \|g_j\|_{L^{np}(X)}$. Using this, the telescoping sum $g_j^{\otimes n} - \varphi_j^{\otimes n} = \sum_{l=1}^n g_j^{\otimes(n-l)} \otimes (g_j - \varphi_j) \otimes \varphi_j^{\otimes(l-1)}$, the triangle inequality, and Lemma 3.6, it follows that

$$\begin{aligned}
\left\| g_j^{\otimes n} - \varphi_j^{\otimes n} \right\|_{L^{np}(X)^{\otimes n}} &\leq \sum_{l=1}^n \left\| g_j^{\otimes(n-l)} \otimes (g_j - \varphi_j) \otimes g_j^{\otimes(l-1)} \right\|_{L^{np}(X)^{\otimes n}} \\
&\leq \sum_{l=1}^n \|g_j\|_{L^{np}(X)}^{n-l} \|g_j - \varphi_j\|_{L^{np}(X)} \|\varphi_j\|_{L^{np}(X)}^{l-1} \\
&\leq n \left(1 + \|g_j\|_{L^{np}(X)}\right)^{n-1} \|g_j - \varphi_j\|_{L^{np}(X)} < \frac{\varepsilon}{2m}.
\end{aligned} \tag{51}$$

Hence, by using (50) and (51), we conclude for $\varphi := \sum_{j=1}^m \varphi_j^{\otimes n} \in \mathcal{NN}_{d,n}^\rho$ that

$$\|g - \varphi\|_{L^{np}(X)^{\otimes n}} \leq \left\| g - \sum_{j=1}^m g_j^{\otimes n} \right\|_{L^{np}(X)^{\otimes n}} + \sum_{j=1}^m \left\| g_j^{\otimes n} - \varphi_j^{\otimes n} \right\|_{L^{np}(X)^{\otimes n}} < \frac{\varepsilon}{2} + \sum_{j=1}^m \frac{\varepsilon}{2m} = \varepsilon.$$

Since $\varepsilon > 0$ and $g \in L^{np}_{\text{diag}}(X)^{\otimes n}$ were chosen arbitrarily, $\mathcal{NN}_{d,n}^\rho$ is dense in $L^{np}_{\text{diag}}(X)^{\otimes n}$. $\square$

Proof of Theorem 5.4. Fix some $G \in L^p(\Omega, \mathcal{F}_T, \mathbb{P})$ and $\varepsilon > 0$. Then, by using Theorem 4.2, there exists some $N \in \mathbb{N}$ and $g_n \in L^{np}_{\text{diag}}(X)^{\otimes n}$, $n = 0, ..., N$, such that

$$\left\| G - \sum_{n=0}^N J_n^\circ(g_n)_T \right\|_{L^p(\mathbb{P})} < \frac{\varepsilon}{2}. \tag{52}$$

Now, for every $n = 1, ..., N$, we apply Proposition 5.3 to obtain some $\varphi_n \in \mathcal{NN}^\rho_{d,n}$ such that

$$\|g_n - \varphi_n\|_{L^{np}(X)^{\otimes n}} < \frac{\varepsilon}{2C_{n,p}N}, \tag{53}$$

where the constant $C_{n,p} > 0$ was introduced in Lemma 3.9 (iii). Hence, by using $\varphi_0 := g_0 \in \mathbb{R}$, Minkowksi's inequality, Lemma 3.9 (iii), and the inequalities (52)+(53), it follows that

$$\begin{aligned}
\left\| G - \sum_{n=0}^N J_n^\circ(\varphi_n)_T \right\|_{L^p(\mathbb{P})} &\leq \left\| G - \sum_{n=0}^N J_n^\circ(g_n)_T \right\|_{L^p(\mathbb{P})} + \left\| \sum_{n=0}^N J_n^\circ(g_n - \varphi_n)_T \right\|_{L^p(\mathbb{P})} \\
&\leq \left\| G - \sum_{n=0}^N J_n^\circ(g_n)_T \right\|_{L^p(\mathbb{P})} + \sum_{n=1}^N \| J_n^\circ(g_n - \varphi_n)_T \|_{L^p(\mathbb{P})} \\
&\leq \left\| G - \sum_{n=0}^N J_n^\circ(g_n)_T \right\|_{L^p(\mathbb{P})} + \sum_{n=1}^N C_{n,p} \|g_n - \varphi_n\|_{L^{np}(X)^{\otimes n}} \\
&< \frac{\varepsilon}{2} + \sum_{n=1}^N C_{n,p} \frac{\varepsilon}{2C_{n,p}N} = \varepsilon,
\end{aligned}$$

which completes the proof. $\square$

7.4.2. *Proof of results in Section 5.2.*

Lemma 7.6. *Let X be a continuous semimartingale satisfying Assumption 2.1, let $p, r \in [1, \infty)$, $n \in \mathbb{N}$, let $\rho \in C(\mathbb{R})$, and let Assumption 5.5 hold. Then, $\mathcal{RN}^\rho_{d,n} \subseteq L^r(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}}; \overline{L^{np}(X)^{\otimes n}})$.*

Proof. First, we observe that $(\overline{L^{np}(X)^{\otimes n}}, \|\cdot\|_{L^{np}(X)^{\otimes n}})$ is as completion of $L^{np}(X)^{\otimes n}$ a Banach space, which we claim to be separable. To this end, we use the polynomial functions $\text{Pol}_{\mathbb{Q}}([0,T];\mathbb{R}^d) := \left\{ [0,T] \ni t \mapsto \sum_{n=0}^N q_n t^n \in \mathbb{R}^d : N \in \mathbb{N}_0, \, q_1, ..., q_N \in \mathbb{Q}^d \right\} \subseteq D_l([0,T];\mathbb{R}^d)$ to define the countable set

$$\text{Pol}_{\mathbb{Q}}([0,T];\mathbb{R}^d)^{\otimes n} := \left\{ p_1 \otimes \cdots \otimes p_n \in L^{np}(X)^{\otimes n} : p_1, ..., p_n \in \text{Pol}_{\mathbb{Q}}([0,T];\mathbb{R}^d) \right\}.$$

Then, by applying the Stone-Weierstrass theorem componentwise, it follows that $\text{Pol}_{\mathbb{Q}}([0,T];\mathbb{R}^d)$ is dense in $C([0,T];\mathbb{R}^d)$ with respect to $\|\cdot\|_\infty$. Hence, by combining Lemma 3.2 (ii) with Lemma 7.4 (as in the proof of Lemma 7.5), we conclude that $\text{Pol}_{\mathbb{Q}}([0,T];\mathbb{R}^d)$ is dense in $L^p(X)$. Thus, by similar arguments as in the proof of Proposition 5.3, it follows that $\text{Pol}_{\mathbb{Q}}([0,T];\mathbb{R}^d)^{\otimes n}$ is dense in $L^{np}(X)^{\otimes n}$.

Next, we show that every $\widetilde{\varphi} \in \mathcal{RN}^\rho_{d,n}$ is $\widetilde{\mathbb{P}}$-strongly measurable. By linearity of $\mathcal{RN}^\rho_{d,n}$, it suffices to prove for every fixed $j \in \mathbb{N}$ as well as for every fixed $\widetilde{\mathcal{F}}_{\text{rand}}/\mathcal{B}(\mathbb{R})$-measurable and $\widetilde{\mathbb{P}}$-a.s. bounded random variable $\widetilde{y}_n : \widetilde{\Omega} \to \mathbb{R}$ that

$$\widetilde{\Omega} \ni \widetilde{\omega} \quad \mapsto \quad \widetilde{R}_j(\widetilde{\omega}) := \widetilde{y}_j(\omega)\rho\left(\widetilde{a}_j(\omega) \cdot + \widetilde{b}_j(\omega)\right)^{\otimes n} \in L^{np}(X)^{\otimes n} \subseteq \overline{L^{np}(X)^{\otimes n}} \tag{54}$$

is $\widetilde{\mathbb{P}}$-strongly measurable, where $\widetilde{y}_j(\omega)\rho\big(\widetilde{a}_j(\omega) \cdot + \widetilde{b}_j(\omega)\big)$ denotes $[0,T] \ni t \mapsto \widetilde{y}_j(\omega)\rho\big(\widetilde{a}_j(\omega)t + \widetilde{b}_j(\omega)\big) \in \mathbb{R}^d$ with $\rho \in C(\mathbb{R})$ applied componentwise. To this end, we first show for every fixed $y \in \mathbb{R}$ that the map

$$\mathbb{R}^d \times \mathbb{R}^d \ni (a,b) \quad \mapsto \quad y\rho(a \cdot + b) \in L^{np}(X) \tag{55}$$

is continuous. Indeed, fix some $\varepsilon > 0$ and let $(a_k, b_k)_{k \in \mathbb{N}} \subseteq \mathbb{R}^d \times \mathbb{R}^d$ be an arbitrary sequence converging to any fixed $(a,b) \in \mathbb{R}^d \times \mathbb{R}^d$, where we use the notation $a_k := (a_{k,1}, ..., a_{k,d})^\top \in \mathbb{R}^d$ and $a := (a_1, ..., a_d)^\top \in \mathbb{R}^d$, and analogously for the vectors $b_k, b \in \mathbb{R}^d$. Then, we can define the finite constants $C_a := \|a\| + \sup_{k \in \mathbb{N}} \|a_k\| \geq 0$ and $C_b := \|b\| + \sup_{k \in \mathbb{N}} \|b_k\| \geq 0$ as well as the compact set $K := \{yt + z : t \in [0,T], \, y \in [-C_a, C_a], \, z \in [-C_b, C_b]\}$. Hence, by using that $\rho \in C(\mathbb{R})$

is continuous, thus uniformly continuous on the compact subset $K \subseteq \mathbb{R}$, there exists some $\delta > 0$ such that for every $s_1, s_2 \in K$ with $|s_1 - s_2| < \delta$ it holds that

$$|\rho(s_1) - \rho(s_2)| < \frac{\varepsilon}{(1+|y|)C_{np,T,X}\sqrt{d}}, \tag{56}$$

where $C_{np,T,X} > 0$ is the constant introduced in Lemma 3.2 (ii). Now, by using that $(a_k, b_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}^d \times \mathbb{R}^d$ converges to $(a,b) \in \mathbb{R}^d \times \mathbb{R}^d$, there exists some $k_0 \in \mathbb{N}$ such that for every $k \in \mathbb{N} \cap [k_0, \infty)$ it holds that $\|(a,b) - (a_k, b_k)\| < \delta/(1+T)$. This implies for every $i = 1, ..., d$ and $t \in [0,T]$ that

$$\left|(a_i t + b_i) - (a_{k,i} t + b_{k,i})\right| \leq |a_i - a_{k,i}|T + |b_i - b_{k,i}| \leq (1+T)\|(a,b) - (a_k,b_k)\| < \delta. \tag{57}$$

Thus, by using Lemma 3.2 (ii) and by combining (56) with (57), it follows for every $k \in \mathbb{N} \cap [k_0, \infty)$ that

$$\begin{aligned}
\|y\rho(a\cdot+b) - y\rho(a_k\cdot+b_k)\|_{L^{np}(X)} &\leq C_{np,T,X}|y|\|\rho(a\cdot+b) - \rho(a_k\cdot+b_k)\|_\infty \\
&\leq |y| C_{np,T,X} \sup_{t\in[0,T]} \left(\sum_{i=1}^d |\rho(a_i t + b_i) - \rho(a_{k,i}t + b_{k,i})|^2\right)^{\frac{1}{2}} \\
&\leq |y| C_{np,T,X}\sqrt{d} \max_{i=1,\dots,d} \sup_{t\in[0,T]} |\rho(a_i t + b_i) - \rho(a_{k,i}t + b_{k,i})| \\
&< |y| C_{np,T,X}\sqrt{d} \frac{\varepsilon}{(1+|y|)C_{np,T,X}\sqrt{d}} = \varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ was chosen arbitrarily, this shows that (55) is continuous. Now, we use this to conclude that

$$\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \ni (a,b,y) \quad \mapsto \quad y\rho(a\cdot+b)^{\otimes n} \in L^{np}(X)^{\otimes n} \tag{58}$$

is also continuous. To this end, let $(a_k, b_k, y_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}$ be an arbitrary sequence converging to any fixed $(a,b,y) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}$. Then, by using the triangle inequality, the telescoping sum $\rho(a\cdot+b)^{\otimes n} - \rho(a_k\cdot+b_k)^{\otimes n} = \sum_{l=1}^n \rho(a\cdot+b)^{\otimes(n-l)} \otimes (\rho(a\cdot+b) - \rho(a_k\cdot+b_k)) \otimes \rho(a_k\cdot+b_k)^{\otimes(l-1)}$, Lemma 3.6, and that (55) is continuous, it follows that

$$\begin{aligned}
&\lim_{k\to\infty} \left\|y\rho(a\cdot+b)^{\otimes n} - y_k\rho(a_k\cdot+b_k)^{\otimes n}\right\|_{L^{np}(X)^{\otimes n}} \\
&\quad\leq \lim_{k\to\infty} |y - y_k| \left\|\rho(a\cdot+b)^{\otimes n}\right\|_{L^{np}(X)^{\otimes n}} + \lim_{k\to\infty} |y_k| \left\|\rho(a\cdot+b)^{\otimes n} - \rho(a_k\cdot+b_k)^{\otimes n}\right\|_{L^{np}(X)^{\otimes n}} \\
&\quad\leq |y| \lim_{k\to\infty} \sum_{l=1}^n \left\|\rho(a\cdot+b)^{\otimes(n-l)} \otimes (\rho(a\cdot+b) - \rho(a_k\cdot+b_k)) \otimes \rho(a_k\cdot+b_k)^{\otimes(l-1)}\right\|_{L^{np}(X)^{\otimes n}} \\
&\quad= |y| \lim_{k\to\infty} \sum_{l=1}^n \|\rho(a\cdot+b)\|_{L^{np}(X)}^{n-l} \|\rho(a\cdot+b) - \rho(a_k\cdot+b_k)\|_{L^{np}(X)} \|\rho(a_k\cdot+b_k)\|_{L^{np}(X)}^{l-1} \\
&\quad= n|y|\|\rho(a\cdot+b)\|_{L^{np}(X)}^{n-1} \lim_{k\to\infty} \|\rho(a\cdot+b) - \rho(a_k\cdot+b_k)\|_{L^{np}(X)} = 0,
\end{aligned}$$

which shows that (58) is indeed continuous. Hence, the map (54) is measurable as composition of the $\widetilde{\mathcal{F}}/\mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R})$-measurable map $\widetilde{\Omega} \ni \widetilde{\omega} \mapsto (\widetilde{a}_j(\widetilde{\omega}), \widetilde{b}_j(\widetilde{\omega}), \widetilde{y}_j(\widetilde{\omega})) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}$ and the continuous map (58). Since $(\overline{L^{np}(X)^{\otimes n}}, \|\cdot\|_{L^{np}(X)^{\otimes n}})$ is by the first step separable, it follows from [31, Corollary 1.1.10 & Proposition 1.1.16.(2)] that the map (54) is $\widetilde{\mathbb{P}}$-strongly measurable.

Finally, we show that $\mathcal{RN}_{d,n}^\rho \subseteq L^r(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}}; \overline{L^{np}(X)^{\otimes n}})$. Again, by linearity of $\mathcal{RN}_{d,n}^\rho$, it suffices to prove that $\widetilde{R}_j \in L^r(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}}; \overline{L^{np}(X)^{\otimes n}})$, where $\widetilde{R}_j : \widetilde{\Omega} \to \overline{L^{np}(X)^{\otimes n}}$ is defined in (54). Indeed, by using that $\widetilde{y}_j : \widetilde{\Omega} \to \mathbb{R}$ is $\widetilde{\mathbb{P}}$-a.s. bounded, Lemma 3.6 together with the fact that $(\widetilde{a}_j, \widetilde{b}_j) \sim (\widetilde{a}_1, \widetilde{b}_1)$ are identically distributed, Lemma 3.2 (ii), and Assumption 5.5, it follows that

$$\begin{aligned}
\widetilde{\mathbb{E}}\left[\left\|\widetilde{R}_j\right\|_{L^{np}(X)^{\otimes n}}^r\right] &= \widetilde{\mathbb{E}}\left[\left\|\widetilde{y}_j \rho\left(\widetilde{a}_j \cdot + \widetilde{b}_j\right)^{\otimes n}\right\|_{L^{np}(X)^{\otimes n}}^r\right] \\
&\leq \|\widetilde{y}_j\|_{L^\infty(\widetilde{\mathbb{P}})}^r \widetilde{\mathbb{E}}\left[\left\|\rho\left(\widetilde{a}_j \cdot + \widetilde{b}_j\right)^{\otimes n}\right\|_{L^{np}(X)^{\otimes n}}^r\right] \\
&= \|\widetilde{y}_j\|_{L^\infty(\widetilde{\mathbb{P}})}^r \widetilde{\mathbb{E}}\left[\left\|\rho\left(\widetilde{a}_1 \cdot + \widetilde{b}_1\right)\right\|_{L^{np}(X)}^{nr}\right] \\
&\leq C_{np,T,X} \|\widetilde{y}_j\|_{L^\infty(\widetilde{\mathbb{P}})}^r \widetilde{\mathbb{E}}\left[\left\|\rho\left(\widetilde{a}_1 \cdot + \widetilde{b}_1\right)\right\|_\infty^{nr}\right] < \infty.
\end{aligned}$$

This shows that $\widetilde{R}_j \in L^r(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}}; \overline{L^{np}(X)^{\otimes n}})$ and thus $\mathcal{RN}_{d,n}^\rho \subseteq L^r(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}}; \overline{L^{np}(X)^{\otimes n}})$. □

Lemma 7.7. *Let X be a continuous semimartingale satisfying Assumption 2.1, let $p, r \in [1,\infty)$, $g \in L^p(X)$, let $\rho \in C(\mathbb{R})$ be non-polynomial, and let Assumption 5.5 hold. Then, for every $\varepsilon > 0$ there exists some $\widetilde{\varphi} \in \mathcal{RN}_{d,1}^\rho$ such that $\widetilde{\mathbb{E}}\big[\|g - \widetilde{\varphi}\|_{L^p(X)}^r\big]^{1/r} < \varepsilon$.*

Proof. For $p, r \in [1,\infty)$, let $g \in L^p(X)$ and fix some $\varepsilon > 0$. Since $\rho \in C(\mathbb{R})$ is non-polynomial, there exists by Lemma 7.5 a fully trained neural network $\varphi := \sum_{n=1}^N y_n \rho(a_n \cdot + b_n) \in \mathcal{NN}_{d,1}^\rho$, with some $a_n := (a_{n,1}, ..., a_{n,d})^\top \in \mathbb{R}^d$, $b_n := (b_{n,1}, ..., b_{n,d})^\top \in \mathbb{R}^d$, and $y_n \in \mathbb{R}$, $n = 1, ..., N$, such that

$$\|g - \varphi\|_{L^{np}(X)} < \frac{\varepsilon}{3}. \tag{59}$$

Hereby, $y_n \rho(a_n \cdot + b_n)$ denotes the function $[0,T] \ni t \mapsto y_n \rho(a_n t + b_n) \in \mathbb{R}^d$ with $\rho \in C(\mathbb{R})$ being applied componentwise. Now, for every fixed $n = 1, ..., N$ and $m, j \in \mathbb{N}$, we define the map

$$\widetilde{\Omega} \ni \widetilde{\omega} \quad \mapsto \quad \widetilde{R}_{n,m,j}(\widetilde{\omega}) := \widetilde{y}_{n,m,j}(\widetilde{\omega}) \rho\left(\widetilde{a}_j(\widetilde{\omega}) \cdot + \widetilde{b}_j(\widetilde{\omega})\right) \in L^p(X) \subseteq \overline{L^p(X)}, \tag{60}$$

where $\rho\big(\widetilde{a}_j(\widetilde{\omega}) \cdot + \widetilde{b}_j(\widetilde{\omega})\big)$ denotes the function $[0,T] \ni t \mapsto \rho\big(\widetilde{a}_j(\omega) t + \widetilde{b}_j(\omega)\big) \in \mathbb{R}^d$ and where

$$\widetilde{\Omega} \ni \widetilde{\omega} \quad \mapsto \quad \widetilde{y}_{n,m,j}(\widetilde{\omega}) := \frac{y_n}{C_{n,m}} \mathbb{1}_{E_{n,m}}(\widetilde{a}_j(\widetilde{\omega}), \widetilde{b}_j(\widetilde{\omega})) \in \mathbb{R}$$

with $E_{n,m} := \big\{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d : \|(x,y) - (a_n, b_n)\| < \frac{1}{m}\big\} \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d)$ and $C_{n,m} := \widetilde{\mathbb{P}}\big[\big\{\widetilde{\omega} \in \widetilde{\Omega} : (\widetilde{a}_j(\widetilde{\omega}), \widetilde{b}_j(\widetilde{\omega})) \in E_{n,m}\big\}\big] > 0$ due to Assumption 5.5. Moreover, by using that the random variable $\widetilde{y}_{n,m,j} : \widetilde{\Omega} \to \mathbb{R}$ is $\widetilde{\mathcal{F}}_{\text{rand}}/\mathcal{B}(\mathbb{R})$-measurable and $\widetilde{\mathbb{P}}$-a.s. bounded, we have $\widetilde{R}_{n,m,j} \in \mathcal{RN}_{d,1}^\rho$ and thus $\widetilde{R}_{n,m,j} \in L^r(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}}; \overline{L^p(X)})$ by Lemma 7.6.

Now, we show that there exists some $m_0 \in \mathbb{N}$ such that for every fixed $n = 1, ..., N$ the expectation $\widetilde{\mathbb{E}}\big[\widetilde{R}_{n,m_0,1}\big] \in \overline{L^p(X)}$ is $\frac{\varepsilon}{3N}$-close to the function $y_n \rho(a_n \cdot + b_n) \in \overline{L^p(X)}$ with respect to $\|\cdot\|_{L^p(X)}$. To this end, we first observe that $E_{n,m+1} \subseteq E_{n,m} \subseteq E_{n,1}$ holds true, for all $m \in \mathbb{N}$, and that the set $K := \{xt + z : t \in [0,T],\ (x,z) \in E_{n,1}\}$ is compact. Hence, by using that $\rho \in C(\mathbb{R})$ is continuous, thus uniformly continuous on the compact subset $K \subseteq \mathbb{R}$, there exists some $\delta > 0$ such that for every $s_1, s_2 \in K$ with $|s_1 - s_2| < \delta$ it holds that

$$|\rho(s_1) - \rho(s_2)| < \frac{\varepsilon}{3N(1 + |y_n|) C_{p,T,X} \sqrt{d}}, \tag{61}$$

where $C_{p,T,X} > 0$ is the constant introduced in Lemma 3.2 (ii). Moreover, by choosing $m_0 \in \mathbb{N}$ large enough such that $m_0 \geq (T+1)/\delta$, we conclude for every $i = 1, ..., d$, $t \in [0,T]$, and $(x,z) \in E_{n,m_0}$ that

$$\big|(a_{n,i} t + b_n) - (x_i t + z_i)\big| \leq |a_{n,i} - x_i| T + |b_{n,i} - z_i| \leq (T+1) \|(a_n, b_n) - (x,y)\| < \frac{T+1}{m_0} \leq \delta, \tag{62}$$

where we use the notation $x := (x_1, ..., x_d)^\top \in \mathbb{R}$ and $z := (z_1, ..., z_d)^\top \in \mathbb{R}$. Hence, by using [31, Proposition 1.2.2], Lemma 3.2 (ii), and the inequalities (61)+(62) together with $(y,z) \in E_{n,m_0}$, it follows that

$$
\begin{aligned}
& \left\| y_n \rho(a_n \cdot + b_n) - \widetilde{\mathbb{E}}\left[\widetilde{R}_{n,m_0,1} \right] \right\|_{L^p(X)} \\
& \quad = \left\| \widetilde{\mathbb{E}}\left[\frac{y_n}{C_{n,m_0}} \mathbb{1}_{E_{n,m_0}}(\widetilde{a}_1, \widetilde{b}_1) \left(\rho(a_n \cdot + b_n) - \rho\left(\widetilde{a}_1 \cdot + \widetilde{b}_1 \right) \right) \right] \right\|_{L^p(X)} \\
& \quad \leq \frac{|y_n|}{C_{n,m_0}} \widetilde{\mathbb{E}}\left[\mathbb{1}_{E_{n,m_0}}(\widetilde{a}_1, \widetilde{b}_1) \left\| \rho(a_n \cdot + b_n) - \rho\left(\widetilde{a}_1 \cdot + \widetilde{b}_1 \right) \right\|_{L^p(X)} \right] \\
& \quad \leq \frac{|y_n|}{C_{n,m_0}} \widetilde{\mathbb{E}}\left[\mathbb{1}_{E_{n,m_0}}(\widetilde{a}_1, \widetilde{b}_1) \right] \sup_{(x,z) \in E_{n,m_0}} \left\| \rho(a_n \cdot + b_n) - \rho(x \cdot + z) \right\|_{L^p(X)} \\
& \quad \leq \frac{|y_n|}{C_{n,m_0}} C_{n,m_0} C_{p,T,X} \sup_{(x,z) \in E_{n,m_0}} \left\| \rho(a_n t + b_n) - \rho(x t + z) \right\|_\infty \\
& \quad = |y_n| C_{p,T,X} \sup_{(x,z) \in E_{n,m_0}} \sup_{t \in [0,T]} \left(\sum_{i=1}^d \left| \rho(a_{n,i} t + b_n) - \rho(x_i t + z_i) \right|^2 \right)^{\frac{1}{2}} \\
& \quad \leq |y_n| C_{p,T,X} \sqrt{d} \max_{i=1,\dots,d} \sup_{(t,(x,z)) \in [0,T] \times E_{n,m_0}} \left| \rho(a_{n,i} t + b_n) - \rho(x_i t + z_i) \right| \\
& \quad < |y_n| C_{p,T,X} \sqrt{d} \frac{\varepsilon}{3N(1 + |y_n|) C_{p,T,X} \sqrt{d}} \leq \frac{\varepsilon}{3N},
\end{aligned} \tag{63}
$$

which shows that $\widetilde{\mathbb{E}}\big[\widetilde{R}_{n,m_0,1}\big] \in \overline{L^p(X)}$ is $\frac{\varepsilon}{3N}$-close to $y_n \rho(a_n \cdot + b_n) \in \overline{L^p(X)}$.

Next, we approximate for every fixed $n = 1, ..., N$ the constant random variable $\big(\widetilde{\omega} \mapsto \widetilde{\mathbb{E}}\big[\widetilde{R}_{n,m_0,1}\big]\big) \in L^1(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}}; \overline{L^p(X)})$ by the average of the i.i.d. sequence $\big(\widetilde{\omega} \mapsto \widetilde{R}_{n,m_0,j}(\widetilde{\omega})\big) \in L^1(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}}; \overline{L^p(X)})$ defined in (60). More precisely, by applying the strong law of large numbers for Banach space-valued random variables in [31, Theorem 3.3.10], we conclude that

$$
\frac{1}{J} \sum_{j=1}^J \widetilde{R}_{n,m_0,j} \quad \overset{J \to \infty}{\longrightarrow} \quad \widetilde{\mathbb{E}}\left[\widetilde{R}_{n,m_0,1} \right] \qquad \text{in} \quad L^1(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}}; \overline{L^p(X)}) \quad \text{and} \quad \widetilde{\mathbb{P}}\text{-a.s..} \tag{64}
$$

In order to generalize the convergence to $L^r(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}}; \overline{L^p(X)})$, we define the sequence of non-negative random variables $(Z_{n,J})_{J \in \mathbb{N}}$ by $Z_{n,J} := \big\| \widetilde{\mathbb{E}}\big[\widetilde{R}_{n,m_0,1}\big] - \frac{1}{J} \sum_{j=1}^J \widetilde{R}_{n,m_0,j} \big\|_{L^p(X)}^r$, $J \in \mathbb{N}$. Then, by using the triangle inequality of $\| \cdot \|_{L^p(X)}$, Minkowski's inequality, [31, Proposition 1.2.2], that $\widetilde{R}_{n,m_0,j} \sim \widetilde{R}_{n,m_0,1}$ is identically distributed, Jensen's inequality, and that $\widetilde{R}_{n,m_0,1} \in \mathcal{RN}_{d,1}^\rho \subseteq L^q(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}}; \overline{L^p(X)})$ for any $q \in [1, \infty)$ (see Lemma 7.6), it follows for every $J \in \mathbb{N}$ and $q \in [1, \infty)$ that

$$
\begin{aligned}
\widetilde{\mathbb{E}}\left[Z_{n,J}^{\frac{q}{r}} \right]^{\frac{1}{q}} & = \widetilde{\mathbb{E}}\left[\left\| \widetilde{\mathbb{E}}\left[\widetilde{R}_{n,m_0,1} \right] - \frac{1}{J} \sum_{j=1}^J \widetilde{R}_{n,m_0,j} \right\|_{L^p(X)}^q \right]^{\frac{1}{q}} \\
& \leq \left\| \widetilde{\mathbb{E}}\left[\widetilde{R}_{n,m_0,1} \right] \right\|_{L^p(X)} + \widetilde{\mathbb{E}}\left[\left(\frac{1}{J} \sum_{j=1}^J \left\| \widetilde{R}_{n,m_0,j} \right\|_{L^p(X)} \right)^q \right]^{\frac{1}{q}} \\
& \leq \widetilde{\mathbb{E}}\left[\left\| \widetilde{R}_{n,m_0,1} \right\|_{L^p(X)} \right] + \frac{1}{J} \sum_{j=1}^J \widetilde{\mathbb{E}}\left[\left\| \widetilde{R}_{n,m_0,j} \right\|_{L^p(X)}^q \right]^{\frac{1}{q}} \\
& \leq 2 \widetilde{\mathbb{E}}\left[\left\| \widetilde{R}_{n,m_0,1} \right\|_{L^p(X)}^q \right]^{\frac{1}{q}} < \infty.
\end{aligned}
$$

Since the right-hand side is finite and does not depend on $J \in \mathbb{N}$, we conclude that $\sup_{J \in \mathbb{N}} \widetilde{\mathbb{E}}\big[Z_{n,J}^{q/r}\big] < \infty$ for all $q \in (1, \infty)$. Hence, the family of random variables $(Z_{n,J})_{J \in \mathbb{N}}$ is by de la Vallée-Poussin's theorem in [38, Theorem 6.19] uniformly integrable. Thus, by using (64), i.e. that $Z_{n,J} \to 0$, $\widetilde{\mathbb{P}}$-a.s., as $J \to \infty$, together with Vitali's convergence theorem, it follows that

$$\lim_{J\to\infty} \mathbb{E}\left[\left\| \widetilde{\mathbb{E}}\left[\widetilde{R}_{n,m_0,1}\right] - \frac{1}{J}\sum_{j=1}^{J} \widetilde{R}_{n,m_0,j} \right\|_{L^p(X)}^r\right] = \lim_{J\to\infty} \widetilde{\mathbb{E}}\left[Z_{n,J}\right] = 0.$$

From this, we conclude that there exists some $J_n \in \mathbb{N}$ such that

$$\widetilde{\mathbb{E}}\left[\left\| \widetilde{\mathbb{E}}\left[\widetilde{R}_{n,m_0,1}\right] - \frac{1}{J_n}\sum_{j=1}^{J_n} \widetilde{R}_{n,m_0,j} \right\|_{L^p(X)}^r\right]^{\frac{1}{r}} < \frac{\varepsilon}{3N}, \tag{65}$$

which shows that the average of the i.i.d. sequence $(\widetilde{R}_{n,m_0,j})_{j=1,...,J_n}$ is $\frac{\varepsilon}{3N}$-close to $\widetilde{\mathbb{E}}\left[\widetilde{R}_{n,m_0,1}\right]$.

Finally, we define the random neural network $\widetilde{\varphi} := \sum_{n=1}^{N} \frac{1}{J_n} \sum_{j=1}^{J_n} \widetilde{R}_{n,m_0,j} \in \mathcal{RN}_{d,1}^{\rho}$. Then, by combining (59), (62), and (65) with the triangle inequality and Minkowski's inequality, it follows that

$$\begin{aligned}
\widetilde{\mathbb{E}}\left[\|g-\widetilde{\varphi}\|_{L^p(X)}^r\right]^{\frac{1}{r}} &\leq \widetilde{\mathbb{E}}\left[\|g-\varphi\|_{L^p(X)}^r\right]^{\frac{1}{r}} + \widetilde{\mathbb{E}}\left[\left\|\varphi - \sum_{n=1}^{N} \widetilde{\mathbb{E}}\left[\widetilde{R}_{n,m_0,1}\right]\right\|_{L^p(X)}^r\right]^{\frac{1}{r}} \\
&\quad + \widetilde{\mathbb{E}}\left[\left\|\sum_{n=1}^{N} \widetilde{\mathbb{E}}\left[\widetilde{R}_{n,m_0,1}\right] - \sum_{n=1}^{N} \frac{1}{J_n}\sum_{j=1}^{J_n} \widetilde{R}_{n,m_0,j}\right\|_{L^p(X)}^r\right]^{\frac{1}{r}} \\
&\leq \frac{\varepsilon}{3} + \sum_{n=1}^{N} \left\|\rho(a_n \cdot + b_n) - \widetilde{\mathbb{E}}\left[\widetilde{R}_{n,m_0,1}\right]\right\|_{L^p(X)} \\
&\quad + \sum_{n=1}^{N} \widetilde{\mathbb{E}}\left[\left\|\widetilde{\mathbb{E}}\left[\widetilde{R}_{n,m_0,1}\right] - \frac{1}{J_n}\sum_{j=1}^{J_n} \widetilde{R}_{n,m_0,j}\right\|_{L^p(X)}^r\right]^{\frac{1}{r}} \\
&\leq \frac{\varepsilon}{3} + N\frac{\varepsilon}{3N} + N\frac{\varepsilon}{3N} = \varepsilon,
\end{aligned}$$

which completes the proof. □

Proof of Proposition 5.8. For $n = 0$ and $p \in [1,\infty)$, the conclusion holds trivially true as $\mathcal{RN}_{d,0}^{\rho} = \mathbb{R}$. Moreover, for $n = 1$, the conclusion follows from Lemma 7.7 by using the identification $L^{np}(X)^{\otimes 1} \cong L^p(X)$. Now, for $n \geq 2$, we fix some $\varepsilon \in (0,1)$ and $g \in L_{\mathrm{diag}}^{np}(X)^{\otimes n}$, which has a representation $g = \sum_{j=1}^{m} g_j^{\otimes n}$, for some $g_1, ..., g_m \in L^{np}(X)$, such that

$$\left\| g - \sum_{j=1}^{m} g_j^{\otimes n} \right\|_{L^{np}(X)^{\otimes n}} < \frac{\varepsilon}{2}. \tag{66}$$

Then, for every fixed $j = 1, ..., m$, we apply Lemma 7.5 to obtain some $\widetilde{\varphi}_j \in \mathcal{RN}_{d,1}^{\rho}$ such that

$$\widetilde{\mathbb{E}}\left[\|g_j - \widetilde{\varphi}_j(\cdot)\|_{L^{np}(X)}^{nr}\right]^{\frac{1}{nr}} < \frac{\varepsilon}{2mn\left(1 + \|g_j\|_{L^{np}(X)}\right)^{n-1}} \leq 1,$$

which implies that $\widetilde{\mathbb{E}}\left[\|\widetilde{\varphi}_j(\cdot)\|_{L^{np}(X)}^{nr}\right]^{nr} \leq 1 + \|g_j\|_{L^{np}(X)}$. Using this, the telescoping sum $g_j^{\otimes n} - \widetilde{\varphi}_j(\widetilde{\omega})^{\otimes n} = \sum_{l=1}^{n} g_j^{\otimes(n-l)} \otimes (g_j - \widetilde{\varphi}_j(\widetilde{\omega})) \otimes \widetilde{\varphi}_j(\widetilde{\omega})^{\otimes(l-1)}$, Minkowski's inequality, Lemma 3.6, Hölder's inequality, and Jensen's inequality, it follows that

$$
\begin{aligned}
\widetilde{\mathbb{E}}\left[\left\|g_j^{\otimes n}-\widetilde{\varphi}_j(\cdot)^{\otimes n}\right\|_{L^{np}(X)^{\otimes n}}^r\right]^{\frac{1}{r}} &\leq \sum_{l=1}^n \widetilde{\mathbb{E}}\left[\left\|g_j^{\otimes(n-l)} \otimes\left(g_j-\widetilde{\varphi}_j(\cdot)\right) \otimes \widetilde{\varphi}_j(\cdot)^{\otimes(l-1)}\right\|_{L^{np}(X)^{\otimes n}}^r\right]^{\frac{1}{r}} \\
&\leq \sum_{l=1}^n\left\|g_j\right\|_{L^{np}(X)}^{n-l} \widetilde{\mathbb{E}}\left[\left\|g_j-\widetilde{\varphi}_j(\cdot)\right\|_{L^{np}(X)}^r\left\|\widetilde{\varphi}_j(\cdot)\right\|_{L^{np}(X)}^{(l-1) r}\right]^{\frac{1}{r}} \\
&\leq \sum_{l=1}^n\left\|g_j\right\|_{L^{np}(X)}^{n-l} \widetilde{\mathbb{E}}\left[\left\|g_j-\widetilde{\varphi}_j(\cdot)\right\|_{L^{np}(X)}^{lr}\right]^{\frac{1}{lr}} \widetilde{\mathbb{E}}\left[\left\|\widetilde{\varphi}_j(\cdot)\right\|_{L^{np}(X)}^{lr}\right]^{\frac{l-1}{lr}} \\
&\leq \widetilde{\mathbb{E}}\left[\left\|g_j-\widetilde{\varphi}_j(\cdot)\right\|_{L^{np}(X)}^{nr}\right]^{\frac{1}{nr}} \sum_{l=1}^n\left\|g_j\right\|_{L^{np}(X)}^{n-l} \widetilde{\mathbb{E}}\left[\left\|\widetilde{\varphi}_j(\cdot)\right\|_{L^{np}(X)}^{nr}\right]^{\frac{l-1}{nr}} \\
&\leq \widetilde{\mathbb{E}}\left[\left\|g_j-\widetilde{\varphi}_j(\cdot)\right\|_{L^{np}(X)}^{nr}\right]^{\frac{1}{nr}} \sum_{l=1}^n\left\|g_j\right\|_{L^{np}(X)}^{n-l}\left(1+\left\|g_j\right\|_{L^{np}(X)}\right)^{l-1} \\
&<\frac{\varepsilon}{2mn\left(1+\left\|g_j\right\|_{L^{np}(X)}\right)^{n-1}} n\left(1+\left\|g_j\right\|_{L^{np}(X)}\right)^{n-1}=\frac{\varepsilon}{2m} .
\end{aligned}
\tag{67}
$$

Hence, by using (66) and (67), we conclude for $\widetilde{\varphi} := \sum_{j=1}^m \widetilde{\varphi}_j^{\otimes n} \in \mathcal{RN}_{d,n}^\rho$ that

$$
\begin{aligned}
\widetilde{\mathbb{E}}\left[\|g-\widetilde{\varphi}(\cdot)\|_{L^{np}(X)^{\otimes n}}^r\right]^{\frac{1}{r}} &\leq \widetilde{\mathbb{E}}\left[\left\|g-\sum_{j=1}^m g_j^{\otimes n}\right\|_{L^{np}(X)^{\otimes n}}^r\right]^{\frac{1}{r}}+\sum_{j=1}^m \widetilde{\mathbb{E}}\left[\left\|g_j^{\otimes n}-\widetilde{\varphi}_j(\cdot)^{\otimes n}\right\|_{L^{np}(X)^{\otimes n}}^r\right]^{\frac{1}{r}} \\
&<\frac{\varepsilon}{2}+\sum_{j=1}^m \frac{\varepsilon}{2m}=\varepsilon,
\end{aligned}
$$

which completes the proof. $\square$

Proof of Theorem 5.9. Fix some $G \in L^p(\Omega, \mathcal{F}_T, \mathbb{P})$ and $\varepsilon > 0$. Then, by using Theorem 4.2, there exists some $N \in \mathbb{N}$ and $g_n \in L^{np}_{\text{diag}}(X)^{\otimes n}$, $n = 0, ..., N$, such that

$$
\left\|G-\sum_{n=0}^N J_n^{\circ}(g_n)_T\right\|_{L^p(\mathbb{P})}<\frac{\varepsilon}{2}. \tag{68}
$$

Now, for every $n = 1, ..., N$, we apply Proposition 5.8 to obtain some $\widetilde{\varphi}_n \in \mathcal{RN}_{d,n}^\rho$ with

$$
\widetilde{\mathbb{E}}\left[\left\|g_n-\widetilde{\varphi}_n(\cdot)\right\|_{L^{np}(X)^{\otimes n}}^r\right]^{\frac{1}{r}}<\frac{\varepsilon}{2C_{n,p}N}, \tag{69}
$$

where the constant $C_{n,p} > 0$ was introduced in Lemma 3.9 (iii). From this, we define the map $\widetilde{\Omega} \ni \widetilde{\omega} \mapsto \sum_{n=0}^N J_n^\circ(\widetilde{\varphi}_n(\widetilde{\omega}))_T \in L^p(\Omega, \mathcal{F}_T, \mathbb{P})$ with $\widetilde{\varphi}_0 := g_0 \in \mathbb{R}$. Since $\widetilde{\Omega} \ni \widetilde{\omega} \mapsto \widetilde{\varphi}_n(\widetilde{\omega}) \in L^{np}(X)^{\otimes n}$ is by Lemma 7.6 $\widetilde{\mathbb{P}}$-strongly measurable and $J_n^\circ(\cdot)_T : L^{np}(X)^{\otimes n} \to L^p(\Omega, \mathcal{F}_T, \mathbb{P})$ is by Lemma 3.9 (iii) linear and bounded, thus continuous and measurable, it follows from [31, Corollary 1.1.11] that $\widetilde{\Omega} \ni \widetilde{\omega} \mapsto \sum_{n=0}^N J_n^\circ(\widetilde{\varphi}_n(\widetilde{\omega}))_T \in L^p(\Omega, \mathcal{F}_T, \mathbb{P})$ is also $\widetilde{\mathbb{P}}$-strongly measurable. Hence, by using Minkowksi's inequality, Lemma 3.9 (iii), and the inequalities (68)+(69), it follows that

$$\begin{aligned}
\widetilde{\mathbb{E}}\left[\left\|G-\sum_{n=0}^{N} J_n^{\circ}(\widetilde{\varphi}_n(\cdot))_T\right\|_{L^p(\mathbb{P})}^r\right]^{\frac{1}{r}} &\leq \widetilde{\mathbb{E}}\left[\left(\left\|G-\sum_{n=0}^{N} J_n^{\circ}(g_n)_T\right\|_{L^p(\mathbb{P})}+\sum_{n=0}^{N}\left\|J_n^{\circ}\left(g_n-\widetilde{\varphi}_n(\cdot)\right)_T\right\|_{L^p(\mathbb{P})}\right)^r\right]^{\frac{1}{r}} \\
&\leq \left\|G-\sum_{n=0}^{N} J_n^{\circ}(g_n)_T\right\|_{L^p(\mathbb{P})}+\sum_{n=1}^{N}\widetilde{\mathbb{E}}\left[\left\|J_n^{\circ}\left(g_n-\widetilde{\varphi}_n(\cdot)\right)_T\right\|_{L^p(\mathbb{P})}^r\right]^{\frac{1}{r}} \\
&\leq \left\|G-\sum_{n=0}^{N} J_n^{\circ}(g_n)_T\right\|_{L^p(\mathbb{P})}+\sum_{n=1}^{N} C_{n,p}\widetilde{\mathbb{E}}\left[\left\|g_n-\widetilde{\varphi}_n(\cdot)\right\|_{L^{np}(X)^{\otimes n}}^r\right]^{\frac{1}{r}} \\
&< \frac{\varepsilon}{2}+\sum_{n=1}^{N} C_{n,p}\frac{\varepsilon}{2C_{n,p}N}=\varepsilon,
\end{aligned}$$

which completes the proof. □

7.4.3. *Proof of results in Section 5.3.*

Proof of Theorem 5.11. Fix $G \in L^p(\Omega,\mathcal{F}_T,\mathbb{P})$ and $\varepsilon > 0$. Then, by Theorem 4.4, there exist some $N \in \mathbb{N}$ and $g_0 \in \mathbb{R}$ as well as $m_n \in \mathbb{N}$ and $g_{n,j,0}, g_{n,j,1} \in L^{np}(X)$, $n = 1,...,N$ and $j = 1,...,m_n$, with

$$\left\|G-g_0-\int_0^T \left(\vartheta_t^{g_{1:N}}\right)^\top dX_t\right\|_{L^p(\mathbb{P})} \leq \frac{\varepsilon}{2}+\inf_{(c,\theta)\in\mathbb{R}\times\Theta^p(X)}\left\|G-c-\int_0^T \theta_t^\top dX_t\right\|_{L^p(\mathbb{P})}, \tag{70}$$

where $\vartheta_t^{g_{1:N}} := \sum_{n=1}^N \sum_{j=1}^{m_n} \frac{W(g_{n,j,0})_t^{n-1}}{(n-1)!} g_{n,j,1}(t)$, for $t \in [0,T]$. Moreover, for every $n = 2,...,N$ and $j = 1,...,m_n$, we follow the proof of Proposition 5.3 to obtain some $\varphi_{n,j,0} \in \mathcal{NN}_{d,1}^{\rho}$ such that

$$\left\|g_{n,j,0}^{\otimes(n-1)}-\varphi_{n,j,0}^{\otimes(n-1)}\right\|_{L^{np}(X)^{\otimes(n-1)}} < \frac{\varepsilon}{4C_{n,p}Nm_n\left(1+\|g_{n,j,1}\|_{L^{np}(X)}\right)}, \tag{71}$$

where $C_{n,p} > 0$ is the constant from Lemma 3.9 (iii). In addition, for every $n = 1,...,N$ and $j = 1,...,m_n$, we apply Lemma 7.5 to obtain some $\varphi_{n,j,1} \in \mathcal{NN}_{d,1}^{\rho}$ such that

$$\|g_{n,j,1}-\varphi_{n,j,1}\|_{L^{np}(X)} < \frac{\varepsilon}{4C_{n,p}Nm_n\left(1+\left\|\varphi_{n,j,0}^{\otimes(n-1)}\right\|_{L^{np}(X)^{\otimes(n-1)}}\right)}, \tag{72}$$

where $\varphi_{n,j,0}^{\otimes 0} := 1$. Hence, by following the proof of Lemma 3.9 (iii) (with Minkowski's inequality, the BDG inequality (with constant $C_p > 0$), and Hölder's inequality), using Lemma 3.9 (iii) with $n-1$ and exponent $\frac{np}{n-1}$, and the inequality (71), it follows for every $n = 2,...,N$ and $j = 1,...,m_n$ that

$$\begin{aligned}
&\mathbb{E}\left[\left|\int_0^T J_{n-1}^{\circ}\left(g_{n,j,0}^{\otimes(n-1)}-\varphi_{n,j,0}^{\otimes(n-1)}\right)_t g_{n,j,1}(t)^\top dX_t\right|^p\right]^{\frac{1}{p}} \\
&\quad\leq \mathbb{E}\left[\sup_{t\in[0,T]}\left|J_{n-1}^{\circ}\left(g_{n,j,0}^{\otimes(n-1)}-\varphi_{n,j,0}^{\otimes(n-1)}\right)_t\right|^{\frac{np}{n-1}}\right]^{\frac{n-1}{np}}\mathbb{E}\left[\left(\int_0^T\left|g_{n,j,1}(t)^\top dA_t\right|\right)^{np}\right]^{\frac{1}{np}} \\
&\quad\quad+C_p\mathbb{E}\left[\sup_{t\in[0,T]}\left|J_{n-1}^{\circ}\left(g_{n,j,0}^{\otimes(n-1)}-\varphi_{n,j,0}^{\otimes(n-1)}\right)_t\right|^{\frac{np}{n-1}}\right]^{\frac{n-1}{np}}\mathbb{E}\left[\left(\int_0^T g_{n,j,1}(t)^\top d\langle M\rangle_t g_{n,j,1}(t)\right)^{\frac{np}{2}}\right]^{\frac{1}{np}} \\
&\quad\leq \max(1,C_p)C_{n-1,np/(n-1)}\left\|g_{n,j,0}^{\otimes(n-1)}-\varphi_{n,j,0}^{\otimes(n-1)}\right\|_{L^{np}(X)^{\otimes(n-1)}}\|g_{n,j,1}\|_{L^{np}(X)} \\
&\quad< C_{n,p}\frac{\varepsilon}{4C_{n,p}Nm_n\left(1+\|g_{n,j,1}\|_{L^{np}(X)}\right)}\|g_{n,j,1}\|_{L^{np}(X)} \leq \frac{\varepsilon}{4Nm_n}.
\end{aligned} \tag{73}$$

Furthermore, by using similar arguments as in (73) and by inserting the inequality (72), we conclude for every $n = 2, ..., N$ and $j = 1, ..., m_n$ that

$$
\begin{aligned}
&\mathbb{E}\left[\left|\int_0^T J_{n-1}^\circ\left(\varphi_{n,j,0}^{\otimes(n-1)}\right)_t \left(g_{n,j,1}(t) - \varphi_{n,j,1}(t)\right)^\top dX_t\right|^p\right]^{\frac{1}{p}} \\
&\quad \leq \mathbb{E}\left[\sup_{t\in[0,T]}\left|J_{n-1}^\circ\left(\varphi_{n,j,0}^{\otimes(n-1)}\right)_t\right|^{\frac{np}{n-1}}\right]^{\frac{n-1}{np}} \mathbb{E}\left[\left|\int_0^T \left(g_{n,j,1}(t) - \varphi_{n,j,1}(t)\right)^\top dA_t\right|^{np}\right]^{\frac{1}{np}} \\
&\qquad + C_p\mathbb{E}\left[\sup_{t\in[0,T]}\left|J_{n-1}^\circ\left(\varphi_{n,j,0}^{\otimes(n-1)}\right)_t\right|^{\frac{np}{n-1}}\right]^{\frac{n-1}{np}} \mathbb{E}\left[\left|\int_0^T \left(g_{n,j,1}(t) - \varphi_{n,j,1}(t)\right)^\top dM_t\right|^{np}\right]^{\frac{1}{np}} \\
&\quad \leq \max(1, C_p) C_{n-1,np/(n-1)} \left\|\varphi_{n,j,0}^{\otimes(n-1)}\right\|_{L^{np}(X)^{\otimes(n-1)}} \left\|g_{n,j,1} - \varphi_{n,j,1}\right\|_{L^{np}(X)} \\
&\quad < C_{n,p}\left\|\varphi_{n,j,0}^{\otimes(n-1)}\right\|_{L^{np}(X)^{\otimes(n-1)}} \frac{\varepsilon}{4C_{n,p}Nm_n\left(1 + \left\|\varphi_{n,j,0}^{\otimes(n-1)}\right\|_{L^{np}(X)^{\otimes(n-1)}}\right)} \leq \frac{\varepsilon}{4Nm_n}.
\end{aligned}
\tag{74}
$$

Thus, by using $\varphi_0 := g_0 \in \mathbb{R}$, the inequality (70), Minkowski's inequality together with Proposition 3.10, and Lemma 3.5 (if $n = 1$) or the inequalities (73)+(74) (if $n \geq 2$), it follows that

$$
\begin{aligned}
&\left\|G - \varphi_0 - \int_0^T \left(\vartheta_t^{\varphi_{1:N}}\right)^\top dX_t\right\|_{L^p(\mathbb{P})} - \inf_{(c,\theta)\in\mathbb{R}\times\Theta^p(X)} \left\|G - c - \int_0^T \theta_t^\top dX_t\right\|_{L^p(\mathbb{P})} \\
&\leq \left\|\int_0^T \left(\vartheta_t^{g_{1:N}}\right)^\top dX_t - \int_0^T \left(\vartheta_t^{\varphi_{1:N}}\right)^\top dX_t\right\|_{L^p(\mathbb{P})} \\
&\quad + \left\|G - g_0 - \int_0^T \left(\vartheta_t^{g_{1:N}}\right)^\top dX_t\right\|_{L^p(\mathbb{P})} - \inf_{(c,\theta)\in\mathbb{R}\times\Theta^p(X)} \left\|G - c - \int_0^T \theta_t^\top dX_t\right\|_{L^p(\mathbb{P})} \\
&\leq \mathbb{E}\left[\left|\sum_{n=1}^N\sum_{j=1}^{m_n}\int_0^T \frac{W(g_{n,j,0})_t^{n-1}}{(n-1)!} g_{n,j,1}(t)^\top dX_t - \sum_{n=1}^N\sum_{j=1}^{m_n}\int_0^T \frac{W(\varphi_{n,j,0})_t^{n-1}}{(n-1)!} \varphi_{n,j,1}(t)^\top dX_t\right|^p\right]^{\frac{1}{p}} + \frac{\varepsilon}{2} \\
&\leq \sum_{n=1}^N\sum_{j=1}^{m_n} \mathbb{E}\left[\left|\int_0^T J_{n-1}^\circ\left(g_{n,j,0}^{\otimes(n-1)} - \varphi_{n,j,0}^{\otimes(n-1)}\right)_t g_{n,j,1}(t)^\top dX_t\right|^p\right]^{\frac{1}{p}} \\
&\quad + \sum_{n=1}^N\sum_{j=1}^{m_n} \mathbb{E}\left[\left|\int_0^T J_{n-1}^\circ\left(\varphi_{n,j,0}^{\otimes(n-1)}\right)_t \left(g_{n,j,1}(t) - \varphi_{n,j,1}(t)\right)^\top dX_t\right|^p\right]^{\frac{1}{p}} + \frac{\varepsilon}{2} \\
&< \sum_{n=1}^N m_n \frac{\varepsilon}{4Nm_n} + \sum_{n=1}^N m_n \frac{\varepsilon}{4Nm_n} + \frac{\varepsilon}{2} = \varepsilon,
\end{aligned}
$$

which completes the proof. □

Proof of Theorem 5.12. Fix $G \in L^p(\Omega, \mathcal{F}_T, \mathbb{P})$ and $\varepsilon > 0$. Then, by Theorem 4.4, there exist some $N \in \mathbb{N}$ and $g_0 \in \mathbb{R}$ as well as $m_n \in \mathbb{N}$ and $g_{n,j,0}, g_{n,j,1} \in L^{np}(X)$, $n = 1, ..., N$ and $j = 1, ..., m_n$, with

$$
\left\|G - g_0 - \int_0^T \left(\vartheta_t^{g_{1:N}}\right)^\top dX_t\right\|_{L^p(\mathbb{P})} \leq \frac{\varepsilon}{2} + \inf_{(c,\theta)\in\mathbb{R}\times\Theta^p(X)} \left\|G - c - \int_0^T \theta_t^\top dX_t\right\|_{L^p(\mathbb{P})}, \tag{75}
$$

where $\vartheta_t^{g_{1:N}} := \sum_{n=1}^N \sum_{j=1}^{m_n} \frac{W(g_{n,j,0})_t^{n-1}}{(n-1)!} g_{n,j,1}(t)$, for $t \in [0,T]$. Moreover, for every $n = 2, ..., N$ and $j = 1, ..., m_n$, we follow the proof of Proposition 5.3 to obtain some $\widetilde{\varphi}_{n,j,0} \in \mathcal{RN}_{d,1}^\rho$ such that

$$
\widetilde{\mathbb{E}}\left[\left\|g_{n,j,0}^{\otimes(n-1)} - \widetilde{\varphi}_{n,j,0}(\cdot)^{\otimes(n-1)}\right\|_{L^{np}(X)^{\otimes(n-1)}}^r\right]^{\frac{1}{r}} < \frac{\varepsilon}{4C_{n,p}Nm_n\left(1 + \|g_{n,j,1}\|_{L^{np}(X)}\right)}, \tag{76}
$$

where $C_{n,p} > 0$ is the constant from Lemma 3.9 (iii). In addition, for every $n = 1, ..., N$ and $j = 1, ..., m_n$, we apply Lemma 7.5 to obtain some $\widetilde{\varphi}_{n,j,1} \in \mathcal{RN}^{\rho}_{d,1}$ such that

$$\widetilde{\mathbb{E}}\left[\|g_{n,j,1} - \widetilde{\varphi}_{n,j,1}(\cdot)\|^{nr}_{L^{np}(X)}\right]^{\frac{1}{nr}} < \frac{\varepsilon}{4C_{n,p}Nm_n\left(1 + \widetilde{\mathbb{E}}\left[\|\widetilde{\varphi}_{n,j,0}(\cdot)^{\otimes(n-1)}\|^{\frac{nr}{n-1}}_{L^{np}(X)^{\otimes(n-1)}}\right]^{\frac{n-1}{nr}}\right)}, \tag{77}$$

where $\widetilde{\varphi}_{n,j,0}(\cdot)^{\otimes 0} := 1$. Hence, we can follow inequality (73) to conclude for every $n = 2, ..., N$ and $j = 1, ..., m_n$ that

$$\begin{aligned} &\widetilde{\mathbb{E}}\left[\left\|\int_0^T J^{\circ}_{n-1}\left(g^{\otimes(n-1)}_{n,j,0} - \widetilde{\varphi}_{n,j,0}(\cdot)^{\otimes(n-1)}\right)_t g_{n,j,1}(t)^\top dX_t\right\|^r_{L^p(\mathbb{P})}\right]^{\frac{1}{r}} \\ &\quad\leq C_{n,p}\widetilde{\mathbb{E}}\left[\left\|g^{\otimes(n-1)}_{n,j,0} - \widetilde{\varphi}_{n,j,0}(\cdot)^{\otimes(n-1)}\right\|^r_{L^{np}(X)^{\otimes(n-1)}}\right]^{\frac{1}{r}} \|g_{n,j,1}\|_{L^{np}(X)} < \frac{\varepsilon}{4Nm_n}. \end{aligned} \tag{78}$$

Furthermore, we can follow the inequality (74) and apply Hölder's inequality to observe for every $n = 2, ..., N$ and $j = 1, ..., m_n$ that

$$\begin{aligned} &\widetilde{\mathbb{E}}\left[\left\|\int_0^T J^{\circ}_{n-1}\left(\widetilde{\varphi}_{n,j,0}(\cdot)^{\otimes(n-1)}\right)_t (g_{n,j,1}(t) - \widetilde{\varphi}_{n,j,1}(\cdot)(t))^\top dX_t\right\|^r_{L^p(\mathbb{P})}\right]^{\frac{1}{p}} \\ &\quad\leq C_{n,p}\widetilde{\mathbb{E}}\left[\left\|\widetilde{\varphi}_{n,j,0}(\cdot)^{\otimes(n-1)}\right\|^r_{L^{np}(X)^{\otimes(n-1)}} \|g_{n,j,1} - \widetilde{\varphi}_{n,j,1}(\cdot)\|^r_{L^{np}(X)}\right]^{\frac{1}{r}} \\ &\quad\leq C_{n,p}\widetilde{\mathbb{E}}\left[\left\|\widetilde{\varphi}_{n,j,0}(\cdot)^{\otimes(n-1)}\right\|^{\frac{nr}{n-1}}_{L^{np}(X)^{\otimes(n-1)}}\right]^{\frac{n-1}{nr}} \widetilde{\mathbb{E}}\left[\|g_{n,j,1} - \widetilde{\varphi}_{n,j,1}(\cdot)\|^{nr}_{L^{np}(X)}\right]^{\frac{1}{nr}} < \frac{\varepsilon}{4Nm_n}. \end{aligned} \tag{79}$$

Thus, by using $\widetilde{\varphi}_0 := g_0 \in \mathbb{R}$, the inequality (75), Minkowski's inequality together with Proposition 3.10, and Lemma 3.5 (if $n = 1$) or the inequalities (78)+(79) (if $n \geq 2$), it follows that

$$\begin{aligned} &\widetilde{\mathbb{E}}\left[\left\|G - \varphi_0 - \int_0^T \left(\vartheta^{\widetilde{\varphi}_{1:N}(\cdot)}_t\right)^\top dX_t\right\|^r_{L^p(\mathbb{P})}\right]^{\frac{1}{r}} - \inf_{(c,\theta)\in\mathbb{R}\times\Theta^p(X)} \left\|G - c - \int_0^T \theta_t^\top dX_t\right\|_{L^p(\mathbb{P})} \\ &\leq \widetilde{\mathbb{E}}\left[\left\|\int_0^T \left(\vartheta^{g_{1:N}}_t\right)^\top dX_t - \int_0^T \left(\vartheta^{\widetilde{\varphi}_{1:N}(\cdot)}_t\right)^\top dX_t\right\|^r_{L^p(\mathbb{P})}\right]^{\frac{1}{r}} \\ &\quad + \left\|G - g_0 - \int_0^T \left(\vartheta^{g_{1:N}}_t\right)^\top dX_t\right\|_{L^p(\mathbb{P})} - \inf_{(c,\theta)\in\mathbb{R}\times\Theta^p(X)} \left\|G - c - \int_0^T \theta_t^\top dX_t\right\|_{L^p(\mathbb{P})} \\ &\leq \widetilde{\mathbb{E}}\left[\left\|\sum_{n=1}^N\sum_{j=1}^{m_n}\int_0^T \frac{W(g_{n,j,0})^{n-1}}{(n-1)!} g_{n,j,1}(t)^\top dX_t - \sum_{n=1}^N\sum_{j=1}^{m_n}\int_0^T \frac{W(\widetilde{\varphi}_{n,j,0}(\cdot))^{n-1}}{(n-1)!} \widetilde{\varphi}_{n,j,1}(\cdot)(t)^\top dX_t\right\|^r_{L^p(\mathbb{P})}\right]^{\frac{1}{r}} \\ &\quad + \frac{\varepsilon}{2} \\ &\leq \sum_{n=1}^N\sum_{j=1}^{m_n}\widetilde{\mathbb{E}}\left[\left\|\int_0^T J^{\circ}_{n-1}\left(g^{\otimes(n-1)}_{n,j,0} - \widetilde{\varphi}_{n,j,0}(\cdot)^{\otimes(n-1)}\right)_t g_{n,j,1}(t)^\top dX_t\right\|^r_{L^p(\mathbb{P})}\right]^{\frac{1}{r}} \\ &\quad + \sum_{n=1}^N\sum_{j=1}^{m_n}\widetilde{\mathbb{E}}\left[\left\|\int_0^T J^{\circ}_{n-1}\left(\widetilde{\varphi}_{n,j,0}(\cdot)^{\otimes(n-1)}_{n,j}\right)_t (g_{n,j,1}(t) - \widetilde{\varphi}_{n,j,1}(\cdot)(t))^\top dX_t\right\|^r_{L^p(\mathbb{P})}\right]^{\frac{1}{r}} + \frac{\varepsilon}{2} \\ &< \sum_{n=1}^N m_n\frac{\varepsilon}{4Nm_n} + \sum_{n=1}^N m_n\frac{\varepsilon}{4Nm_n} + \frac{\varepsilon}{2} = \varepsilon, \end{aligned}$$

which completes the proof. $\square$

REFERENCES

[1] Patrick Billingsley. *Convergence of probability measures*. Probability and statistics section. J. Wiley, New York, 2nd edition, 1999.

[2] Hans Bühler, Lukas Gonon, Josef Teichmann, and Ben Wood. Deep hedging. *Quantitative Finance*, 19(8):1271–1291, 2019.

[3] Robert H. Cameron and William T. Martin. The orthogonal development of non-linear functionals in series of fourier-hermite functionals. *Annals of Mathematics*, 48(2):385–392, 1947.

[4] Peter Carr and Dilip B. Madan. Option valuation using the fast Fourier transform. *Journal of Computational Finance*, 2:61–73, 1999.

[5] Álvaro Cartea, Imanol Pérez Arribas, and Leandro Sánchez-Betancourt. Double-execution strategies using path signatures. *SIAM Journal on Financial Mathematics*, 13(4):1379–1417, 2022.

[6] Kyriakos Chourdakis. Option pricing using the fractional FFT. *Journal of Computational Finance*, 8(2):1–18, 2005.

[7] John Maurice Clark. The representation of functionals of Brownian motion by stochastic integrals. *The Annals of Mathematical Statistics*, 41(4):1282–1295, 1970.

[8] Christa Cuchiero. *Affine and polynomial processes*. PhD thesis, ETH Zurich, 2011.

[9] Christa Cuchiero, Wahid Khosrawi, and Josef Teichmann. A generative adversarial network approach to calibration of local stochastic volatility models. *Risks*, 8(4), 2020.

[10] Christa Cuchiero, Francesca Primavera, and Sara Svaluto-Ferro. Universal approximation theorems for continuous functions of càdlàg paths and Lévy-type signature models. *Finance and Stochastics*, 29:289–342, 2025.

[11] Christa Cuchiero, Philipp Schmocker, and Josef Teichmann. Global universal approximation of functional input maps on weighted spaces. *arXiv preprint 2306.03303*, 2023.

[12] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314, 1989.

[13] José Da Fonseca, Martino Grasselli, and Claudio Tebaldi. Option pricing when correlations are stochastic: an analytical framework. *Review of Derivatives Research*, 10:151–180, 2007.

[14] Claude Dellacherie and Paul-André Meyer. *Probabilities and Potential A*. North Holland, Amsterdam, 1978.

[15] Giulia Di Nunno, Bernt Øksendal, and Frank Proske. *Malliavin Calculus for Lévy Processes with Applications to Finance*. Springer Science + Business Media, Berlin Heidelberg, 2008.

[16] Paolo Di Tella and Hans-Jürgen Engelbert. The chaotic representation property of compensated-covariation stable families of martingales. *Annals of Probability*, 44(6):3965–4005, 11 2016.

[17] Darrell Duffie, Jun Pan, and Kenneth Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376, 2000.

[18] Richard Durrett. *Probability: Theory and examples*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge, 4th edition, 2010.

[19] Stephan Eckstein, Michael Kupper, and Mathias Pohl. Robust risk aggregation with neural networks. *Mathematical Finance*, 30(4):1229–1272, 2020.

[20] Michel Émery. On the Azéma martingales. *Séminaire de probabilités de Strasbourg*, 23:66–87, 1989.

[21] Damir Filipović and Martin Larsson. Polynomial diffusions and applications in finance. *Finance and Stochastics*, 20(4):931–972, 2016.

[22] Gerald B. Folland. *Fourier analysis and its applications*. Brooks/Cole Publishing Company, Belmont, California, 1st edition, 1992.

[23] Lukas Gonon. Random feature neural networks learn Black-Scholes type PDEs without curse of dimensionality. *Journal of Machine Learning Research*, 24(189):1–51, 2023.

[24] Lukas Gonon, Lyudmila Grigoryeva, and Juan-Pablo Ortega. Approximation bounds for random neural networks and reservoir systems. *The Annals of Applied Probability*, 33(1):28–69, 2023.

[25] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[26] Lyudmila Grigoryeva and Juan-Pablo Ortega. Echo state networks are universal. *Neural Networks*, 108:495–508, 2018.

[27] Bruce Hajek. Mean stochastic comparison of diffusions. *Probability Theory and Related Fields*, 68(3):315–329, 1985.

[28] Jiequn Han, Arnulf Jentzen, and Weinan E. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, 115(34):8505–8510, 2018.

[29] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991.

[30] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1):489–501, 2006. Neural Networks.

[31] Tuomas Hytönen, Jan van Neerven, Mark Veraar, and Lutz Weis. *Analysis in Banach Spaces*, volume 63 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge*. Springer, Cham, 2016.

[32] Kiyosi Ito. Multiple Wiener integral. *Journal of the Mathematical Society of Japan*, 3(1):157–169, 05 1951.

[33] Kiyosi Ito. Spectral type of the shift transformation of differential processes with stationary increments. *Transactions of the American Mathematical Society*, 81(2):253–263, 1956.

[34] Farshid Jamshidian. Chaotic expansion of powers and martingale representation. Finance 0506008, University Library of Munich, Germany, Jun 2005.

[35] Ioannis Karatzas and Steven Shreve. *Brownian Motion and Stochastic Calculus*. Springer Science + Business Media, Berlin Heidelberg, 2nd corrected ed. 1998. corr. 6th printing 2004 edition, 1998.

[36] Martin Keller-Ressel. *Affine Processes – Theory and Applications in Finance*. PhD thesis, TU Vienna, 2009.

[37] Franz J. Király and Harald Oberhauser. Kernels for sequentially ordered data. *Journal of Machine Learning Research*, 20(31):1–45, 2019.

[38] Achim Klenke. *Probability Theory: A Comprehensive Course*. Universitext. Springer London, London, 2nd edition, 2014.
[39] Vincent Lacoste. Wiener chaos: a new approach to option hedging. *Mathematical Finance*, 6(2):197–213, 1996.
[40] Jérôme Lelong. Dual pricing of American options by Wiener chaos expansion. *SIAM Journal on Financial Mathematics*, 9(2):493–519, 2018.
[41] Moshe Leshno, Vladimir Ya. Lin, Allan Pinkus, and Shimon Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6):861–867, 1993.
[42] Terry J. Lyons, Michael Caruana, and Thierry Lévy. *Differential Equations Driven by Rough Paths*. Number 1908 in Lecture Notes in Mathematics. Springer Berlin Heidelberg New York, 1998.
[43] Terry J. Lyons, Sina Nejad, and Imanol Perez Arribas. Numerical method for model-free pricing of exotic derivatives in discrete time using rough path signatures. *Applied Mathematical Finance*, 26(6):583–597, 2019.
[44] Terry J. Lyons, Sina Nejad, and Imanol Perez Arribas. Non-parametric pricing and hedging of exotic derivatives. *Applied Mathematical Finance*, 27(6):457–494, 2020.
[45] Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5:115–133, 1943.
[46] Ariel Neufeld and Philipp Schmocker. Universal approximation property of Banach space-valued random feature models including random neural networks. *arXiv preprint arXiv:2312.08410*, 2023.
[47] Ariel Neufeld and Julian Sester. A deep learning approach to data-driven model-free pricing and to martingale optimal transport. *IEEE Transactions on Information Theory*, 69(5):3172–3189, 2023.
[48] Ariel Neufeld, Julian Sester, and Daiying Yin. Detecting data-driven robust statistical arbitrage strategies with deep neural networks. *SIAM Journal on Financial Mathematics*, 15(2):436–472, 2024.
[49] David Nualart. *The Malliavin Calculus and Related Topics*. Probability and Its Applications. Springer Science + Business Media, Berlin Heidelberg, 2nd edition, 2006.
[50] David Nualart and Wim Schoutens. Chaotic and predictable representations for Lévy processes. *Stochastic Processes and their Applications*, 90(1):109 – 122, 2000.
[51] Daniel Ocone. Malliavin's calculus and stochastic integral representations of functional of diffusion processes. *Stochastics*, 12(3-4):161–185, 1984.
[52] Huyên Pham. On quadratic hedging in continuous time. *Mathematical Methods of Operations Research*, 51:315–339, 2000.
[53] Allan Pinkus. Approximation theory of the MLP model in neural networks. *Acta Numerica*, 8:143–195, 1999.
[54] Philip Protter. *Stochastic integration and differential equations*, volume 21 of *Stochastic modelling and applied probability*. Springer, Berlin, 2nd ed., 3rd print edition, 2005.
[55] Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Proceedings of the 20th International Conference on Neural Information Processing Systems*, NIPS'07, pages 1177–1184, Red Hook, NY, USA, 2007. Curran Associates Inc.
[56] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*. Grundlehren der mathematischen Wissenschaften 293. Springer, Berlin, 3rd edition, 1999.
[57] Walter Rudin. *Functional analysis*. International series in pure and applied mathematics. McGraw-Hill, Boston, Mass, 2nd edition, 1991.
[58] Johannes Ruf and Weiguan Wang. Neural networks for option pricing and hedging: a literature review. *Journal of Computational Finance*, 24(1):1–46, 2020.
[59] Raymond A. Ryan. *Introduction to tensor products of Banach spaces*. Springer monographs in mathematics. Springer, London, 2002.
[60] Philipp Schmocker. Universal approximation on path spaces and applications in finance, 2022. University of St.Gallen.
[61] Martin Schweizer. A guided tour through quadratic hedging approaches. SFB 373 Discussion Papers 1999,96, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, 1999.
[62] Justin Sirignano and Rama Cont. Universal features of price formation in financial markets: perspectives from deep learning. *Quantitative Finance*, 19(9):1449–1459, 2019.
[63] Norbert Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60(4):897–936, 1938.
[64] Dongbin Xiu and George Em Karniadakis. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002.

NANYANG TECHNOLOGICAL UNIVERSITY, DIVISION OF MATHEMATICAL SCIENCES, 21 NANYANG LINK, SINGAPORE
Email address: ariel.neufeld@ntu.edu.sg

NANYANG TECHNOLOGICAL UNIVERSITY, DIVISION OF MATHEMATICAL SCIENCES, 21 NANYANG LINK, SINGAPORE
Email address: philippt001@e.ntu.edu.sg