University of Amsterdam

MSc Mathematics

Master Thesis

Deep Learning Algorithms for Partial Stochastic Differential Equations in Hilbert Spaces for flow forwards pricing

Author:
Alvise Sembenico

Supervisor:
dr. Asma Khedher

Examination date:
October 27, 2025

Korteweg-de Vries Institute for
Mathematics

Abstract

We address the numerical pricing of European options on flow forwards when the underlying forward curve evolves as a mild solution of an SPDE in a separable Hilbert function space (Filipović-type), a setting where classical Monte Carlo methods face the curse of dimensionality. We recast the time–t price functional V as the minimizer of a suitable L^2 objective over continuous maps on the forward-curve space and prove that V is Lipschitz in the state under standard semigroup assumptions on the SPDE. Building on a Hilbert-space formulation of feedforward neural networks, we establish a universal approximation result: for discriminatory activations, finite networks are dense in the supremum norm on compact subsets of the state space, hence approximate V arbitrarily well on compacts. We complement this with a comparison to global (noncompact) uniform approximation frameworks in weighted-spaces and finite-dimensional noncompact UAT with asymptotically affine activations. The pricing map for flow forwards is represented via an averaging operator composed with evaluation functionals, and boundedness of these operators on the chosen Hilbert space is verified. To verify numerically the convergence rate, we generate a synthetic dataset for the input and use Monte Carlo method with multi dimensional noise to generate the corresponding label. We then train a finite-dimensional neural network and analyze the converge rate.

The results indicate that Hilbert-space neural networks provide an efficient and theoretically justified surrogate for SPDE-driven term-structure models, bridging rigorous approximation guarantees with practical performance.

Title: Deep Learning Algorithms for Partial Stochastic Differential Equations in Hilbert Spaces for flow forwards pricing
Author: Alvise Sembenico, alvise.sembenico@student.uva.nl, 12380288
Supervisor:
Examination date: October 27, 2025

Korteweg-de Vries Institute for Mathematics
University of Amsterdam
Science Park 105-107, 1098 XG Amsterdam
`http://kdvi.uva.nl`

Contents

Introduction

Curse of dimensionality.

Stochastic differential equations (SDEs) have long been a vibrant research topic, not least because of their prominence in quantitative finance, where they underpin the valuation of derivatives and the management of risk. Closed-form solutions exist only for a limited class of linear or affine models; consequently, substantial effort has been devoted to designing efficient numerical schemes that approximate SDE solutions to a prescribed accuracy. Owing to the *Feynman–Kac* formula, many linear parabolic partial differential equations (PDEs) can be reduced to the problem of simulating appropriate SDEs, so advances in one area immediately benefit the other.

Classical time-stepping schemes such as Euler–Maruyama, Milstein, and higher-order stochastic Taylor methods [42] provide strong or weak convergence at polynomial computational cost *in low dimensions*. Their efficiency, however, deteriorates rapidly once the state dimension grows: the number of Monte Carlo sample paths required to maintain a given error tolerance typically increases *exponentially* with the dimension, a phenomenon referred to as the *curse of dimensionality*. Alternative methods, such as spectral or quasi–Monte Carlo, have been proposed to increase the convergence rate. Quasi–Monte Carlo draws samples from a low-discrepancy sequence instead of pseudo-random numbers; this sequence is designed to place points as uniformly as possible. In some cases, this can lead to significant improvements, raising the convergence rate from $O(N^{-1/2})$ for the standard Monte Carlo method to $O(N^{-1})$; see [34]. However, these refinements alleviate the problem only partially and become impractical when the dimension reaches several dozen, let alone hundreds.

Over the past decade, a rapidly expanding literature has explored neural-network-based solvers as a remedy for this curse [24, 32, 7, 19]. Closely related approaches for forward–backward stochastic differential equations (FBSDEs) have achieved *a posteriori* convergence rates that match, and in some regimes outperform, those obtained by traditional Monte Carlo methods [33, 37, 59, 5, 40, 20, 45, 46]. Looking at application in finance, we analyze neural-network techniques for the *infinite-dimensional* setting and justify them by a universal approximation theorem in the supremum-norm topology on compact sets. Such an extension is indispensable for models whose state evolves in a function space, as is common for the term-structure dynamics of forward contracts.

Motivation.

Many problems in mathematical finance are formulated in infinite-dimensional state spaces. Their dynamics are naturally described by stochastic partial differential equations (SPDEs) driven by a cylindrical or a Q-Wiener processes in a separable Hilbert space $(H, \langle\cdot,\cdot\rangle_H)$. Intuitively, a Q-Wiener process is characterized by its *covariance operator* $Q : H \to H$, which is bounded, self-adjoint, positive, and trace-class. For $0 \le s \le t$ and $h, h' \in H$,

$$\mathbb{E}\big[\langle W_t - W_s, h\rangle_H \,\langle W_t - W_s, h'\rangle_H\big] \;=\; (t-s)\,\langle Qh, h'\rangle_H, \qquad 0 \le s \le t,\; h, h' \in H.$$

The separability of H guarantees the existence of a countable orthonormal basis, while the trace-class condition $\operatorname{Tr} Q < \infty$ ensures that the paths of W take values in H with probability one. Hence, the bilinear form $(h, h') \mapsto \langle Qh, h'\rangle_H$ plays the role of a covariance matrix in finite dimensions, encoding both directionwise variances and correlations.

Standard spatial discretizations, finite differences, Galerkin truncations [6] reduce an SPDE to a high-dimensional SDE, but thereby inherit the same computational bottlenecks.

A case in point is the market for electricity and gas. Here, the fundamental traded objects are *forward contracts* that oblige the seller to deliver a specified amount of energy at a future time (or during a future period) in exchange for a payment today. When the delivery spans a non-trivial interval, we speak of *flow-forwards.* Such contracts are the most liquid instruments on exchanges like the European Energy Exchange (EEX) because energy is typically consumed over extended periods.

Let $[T_1, T_2]$ denote the delivery window, let $t \le \tau \le T_1$ be the current time and the option's exercise date, respectively, and fix a strike $K > 0$. The time-t price of a European call on the flow forward is

$$V(t,\tau) = e^{-r(\tau-t)}\mathbb{E}\Big[\big(\widehat{F}(\tau, T_1, T_2) - K\big)^+ \;\Big|\; \mathcal{F}_t\Big],$$

where r is the (continuously compounded) risk-free rate and

$$\widehat{F}(t, T_1, T_2) \;=\; \frac{1}{T_2 - T_1}\int_{T_1}^{T_2} F(t,T)\,dT$$

is the average of the hypothetical *single-maturity* forward prices $F(t,T)$ over the delivery period. Because such single-maturity forwards are rarely traded for power and gas, their prices $F(t,T)$ must be obtained from a suitable term-structure model rather than directly from the market.

A widely used arbitrage-free framework represents the entire forward curve $T \mapsto F(t,T)$ as the mild solution of an SPDE in an infinite-dimensional Hilbert *function* space (e.g., the Filipović space), which admits continuous evaluation functionals and a right-shift C_0-semigroup [12]. Traditional numerical methods struggle with this setting, whereas neural networks, when formulated on Hilbert spaces, can, in principle, approximate the solution map while keeping only finitely many trainable parameters.

In this thesis we analyze mainly the work in [10, 9], mentioning when necessary alternative approaches. First, we express the pricing functional V as the minimizer of a suitable L^2 objective and, without loss of generality, show that optimization can be restricted to neural networks acting on the underlying Hilbert space. Second, relying on the universal approximation theorem for Hilbert-space neural networks from [9], we prove that V can be approximated arbitrarily well on compact sets whenever the activation function is discriminatory (separating) and the forward-curve dynamics satisfy a Lipschitz condition. It is important to note that there are results concerting non compact-set, for both the finite dimensional case [23], as well as the Hilbert space valued [48]. Numerical experiments confirm that the resulting *finite-dimensional* networks perform competitively with classical Monte Carlo benchmarks.

Structure of the thesis.
The thesis is structured as follows. The first part of the background is dedicated to the construction of the stochastic integral in an infinite-dimensional Hilbert space. To achieve this, we first define the Q-Wiener process and, consequently, martingales in such a space. Equipped with these definitions, we construct the stochastic integral with respect to a Q-Wiener process. Many ideas for the construction are borrowed from standard finite-dimensional, real-valued stochastic integration theory. The stochastic integral is first defined for elementary processes; then, using the Itô isometry and the density of elementary processes in the class of adapted H-valued stochastic processes, we extend the definition to a richer class. Main results, such as Itô's lemma and the Martingale Representation Theorem [28], are presented at the end of the section.

Together with a short section on C_0-semigroup theory, we analyze SPDEs, namely their formulation and their solution types, focusing specifically on the mild solution, whose existence and uniqueness have been established for semilinear SPDEs of the form

$$\begin{cases} dX(t) = \big(AX(t) + F(t,X)\big)\,dt + B(t,X)\,dW_t, \\ X(0) = \xi_0, \end{cases} \tag{0.1}$$

under Lipschitz-type conditions on the coefficients [28]. We then introduce the Filipović function space used throughout, emphasizing continuous evaluations and the right-shift semigroup.

In the second section of the background, we formally introduce neural networks and the universal approximation theorem for Hilbert-space neural networks. The theorem establishes density in the supremum-norm topology on compact sets under assumptions solely on the activation function σ (discriminatory/separating).

In Chapter 3.2 we formulate the flow-forward option-pricing problem on the forward-curve space, prove that the corresponding price functional is Lipschitz and can be approximated in $L^2(\mu)$ for an appropriate measure μ, and, under the separating property of the activation, show that it can be approximated by finite-dimensional neural networks. Finally, we present numerical experiments in which we implement and train such a network and demonstrate convergence.

Notation

- For a topological vector space X and a field $\mathbb{F}$ (generally $\mathbb{R}$ or $\mathbb{C}$), topological dual of X, denoted by $X^\star$, is the space of all continuous linear functionals $f : X \to \mathbb{F}$.
- Let H, K be Hilbert spaces. We denote by $L(H, K)$ the set of bounded linear operators from H to K. If $H = K$, we simply write $L(H)$.
- For a Hilbert space H, $H^\star$ its dual, $\langle \cdot, \cdot \rangle$ denotes the duality pairing between H and $H^\star$, that is, for $x \in H$ and $x^\star \in H^\star$, $\langle x^\star, x \rangle := x^\star(x)$.
- For an operator $A \in L(H)$, the adjoint operator of A is denoted by $A^\star$.
- For X a metric space, $B(X)$ denotes the Borel σ-algebra on X generated by the open balls in X.
- For a Hilbert space $(H, \langle \cdot, \cdot \rangle_H)$ and a subspace $M \subset H$, the notation $M^\perp$ denotes the orthogonal complement of M, defined as $M^\perp = \{h \in H : \langle h, m \rangle_H = 0\, \forall m \in M\}$.
- For Hilbert spaces H, K with the topology induced by the norm, we denote by $C(H, K), C_b(H, K), C_0(H, K)$, and $C_c(H, K)$ respectively the family of continuous functions, bounded functions, functions vanishing at infinity, i.e. $\lim_{\|x\|_H \to \infty} f(x) = 0$, and finally the functions with a compact support, mapping from H to K. Among these, $C_b(H, K)$ and $C_0(H, K)$ are Banach spaces when endowed with the supremum norm, which is written as $\|f\|_{C_b(H,K)}$ and $\|f\|_{C_0(H,K)}$ respectively.
- For $p \in \mathbb{N}^+$, H a Hilbert space with associated norm $\| \cdot \|_H$, $L^p(\Omega, H, \mu)$ is the set of all measurable functions that are L^p integrable with respect to μ, i.e. all $f : \Omega \to H$ such that $\|f\|_{L^p(\Omega,H)} := \left(\int_\Omega \|f(x)\|_H^p \mu(dx) \right)^{\frac{1}{p}} < \infty$. Note that we might omit H and/or μ where no confusion arises. Unless specified, μ denotes the Lebesgue measure. Moreover, for $U \subset \mathbb{R}$, $L^p(U)$ denotes the usual space $L^p(U, B(U), \mathbb{P})$. For two Hilbert spaces H, K and a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we denote by $L^p(\Omega, C_b(H, K))$ the space $L^p(\Omega, C_b(H, K), \mathbb{P})$ where we endowed the space $C_b(H, K)$ with the supremum norm. Formally $L^p(\Omega, C_b(H, K)) := \{f : \Omega \to C_b(H, K) \text{ measurable } : \left(\int_\Omega \sup_{h \in H} \|f(x)(h)\|_K^p \mathbb{P}(dx) \right)^{\frac{1}{p}}\}$, equivalently $L^p(\Omega, C_b(H, K)) := \{f \in \Omega \to C(H, K) \text{ measurable } : \mathbb{E}\left[\|f\|_{C_b(H,K)}^p \right] < \infty\}$.
- For two topological spaces U and V, $C(U, V)$ is the collection of all continuous functions from U to V.

- For an operator $Q : H \to K$, $D(Q)$ is the domain of Q, i.e. $D(Q) \coloneqq \{h \in H : Qh \text{ is well defined}\}$.

- For a function $f : H \to K$ with K endowed with a topology, $\mathrm{supp}(f) \coloneqq \overline{\{u \in U : f(u) \neq 0\}}$, where for a set A, $\overline{A}$ denotes the closure in the topology of H.

- For a function $f : \mathbb{R}^n \to \mathbb{R}^m$ and $x \in \mathbb{R}^n = (x_1, \dots, x_n)$, $\nabla_{x_i} f(x)$ represents the partial derivative of f with respect to x_i.

1 Stochastic Partial Differential Equations

1.1 Semigroup theory

In this section, we give some elementary notions and results concerning semigroup theory. The purpose of this section is to present the essential concepts that will be invoked later in this work.

Definition 1.1.1. *Given $(H, \langle \cdot, \cdot \rangle_H)$ a Hilbert space, a family $S(t) \in L(H)$ for $t \geq 0$ of bounded linear operators on H is called strongly continuous semigroup (C_0 semigroup) if the following properties hold*

1. $S(0) = I$
2. $S(t+s) = S(t)S(s)$ *(Semigroup property),*
3. $\lim_{t \to 0^+} S(t)h = h \quad \forall h \in H$ *(Strong continuity property).*

Example 1.1.2 (Exponentials in a Banach algebra give a C_0-semigroup)**.** *Let $(A, \|\cdot\|)$ be a Banach algebra with unit and fix $a \in A$. Define*

$$e^{ta} := \sum_{n=0}^{\infty} \frac{(ta)^n}{n!} \in A, \; S(t)x := e^{ta}\, x, \quad x \in A, \; t \geq 0.$$

Then $(S(t))_{t \geq 0}$ is a (uniformly continuous) C_0-semigroup on A, in fact, the semigroup property follow directly from the definition and $\|S(t)\| \leq e^{t\|a\|}$.

We present a useful proposition regarding the uniform bound of the C_0-semigroup.

Proposition 1.1.3 (Proposition 1.2 of [50])**.** *For a Hilbert space H and $S(t)$ a C_0-semigroup, for the canonical operator norm $\|\cdot\|_{L(H)}$, there exists an $M \geq 1$ and $\omega \geq 0$ such that*

$$\|S(t)\|_{L(H)} \leq M e^{\omega t} \tag{1.1}$$

for every $t \geq 0$.

Proof. The proof is composed of two steps. The first step consists of proving that $S(t)$ is bounded on $[0, \delta]$ for some $\delta > 0$. Consider the map $t \to \|S(t)\|_{L(H)}$. Suppose, for contradiction, that for every $\delta > 0$, $\|S(t)\|_{L(H)}$ is unbounded on $[0, \delta]$.

Because of this, we can choose the sequence $t_n \in [0, \frac{1}{n}]$ such that $\|S(t_n)\|_{L(H)} \le n$ for every $n \in \mathbb{N}$. Thus, $\sup_n \|S(t_n)\|_{L(H)} = \infty$. By the negation of the Uniform boundedness principle, it must hold that $\exists h \in H$ such that $\sup_{n\in\mathbb{N}} \|S(t_n)h\| = \infty$. However, since $\lim_{n\to\infty} S(t_n)x = x$ because $t_n \to 0$, it must hold that $\lim_{n\to\infty} \|S(t_n)\|_{L(H)} \le 1$, which implies that $\sup_{n\in\mathbb{N}} \|S(t_n)\|_{L(H)} < \infty$, which contradicts the assumption, therefore, there exists a $\delta > 0$ such that the map $t \to \|S(t)\|_{L(H)} < \infty$ for every $0 \le t \le \delta$.

The next step is to extend the result from the previous step to the whole $\mathbb{R}$. Define $M := \sup_{t\in[0,\delta]} \|S(t)\|_{L(H)} < \infty$. Note that $M \ge 1$ since $S(0) = I$. The next series of equalities follows directly from the property of the semigroup operator. For an arbitrary $t = N\delta + r$ where $N \in \mathbb{N}$ and $0 \le r \le \delta$, the following hold true.

$$\begin{aligned}\|S(t)\| = \|S(N\delta + r)\| &= \|S(N\delta)S(r)\| \\ &= \|S(\delta)^N S(r)\| = M^{N+1}.\end{aligned}$$

We can rewrite the above as follows

$$M^{N+1} = MM^N \le MM^{\frac{t}{\delta}} = Me^{\frac{t}{\delta}\log M}. \tag{1.2}$$

Upon choosing $\omega = \frac{t}{\delta}\log M$ we get the desired result. □

In the Proposition above, if $M = 1$, we call $S(t)$ a pseudo-contraction semigroup; if $\alpha = 0$, $S(t)$ is uniformly bounded with respect to t. If $M = 1$ and $\alpha = 0$, $S(t)$ is called the semigroup of contractions, finally, if $M > 1$, then the semi-group is called quasi-contractive.

Moreover, we call S compact if $S(t)$ is compact for every $t \ge 0$. The following Proposition shows the meaning of the strong continuity property.

Proposition 1.1.4 (Section 1.1 of [28]). *For any C_0 semigroup $S(t)$ over a Hilbert space H and for every $h \in H$, the map*

$$t \to S(t)h \tag{1.3}$$

is continuous for $t \in \mathbb{R}^+$.

Proof. Consider a converging sequence in $\mathbb{R}^+ : t_n \to t$. Consider now only the subsequence such that $t_n \ge t$ for every n. It follows

$$\begin{aligned}\|S(t_n)h - S(t)h\|_H &= \|(S(t_n) - S(t))h\|_H \\ &= \|S(t)(S(t_n - t) - I)h\|_H \\ &\le \|S(t)\|_{L(H)}\|S(t_n - t) - I\|_H\|h\|_H,\end{aligned}$$

where we used properties 2 and 3 of Definition 1.1.1. Using property 1 of the same definition, we get that $\|S(t_n - t) - I\|_{L(H)} \to 0$. Using the fact that $S(t) \in L(H)$, the whole expression converges to 0. The exact same argument can be carried out for the part of the sequence where $t_n \le 0$ using $S(t_n)(I - S(t - t_n))$ instead. □

We now move to another fundamental concept when working with C_0 semigroup.

Definition 1.1.5. *Let $S(t)$ be a C_0 semigroup on a Hilbert space H, define $D(A) \subseteq H$*

$$D(A) = \{h \in H : \lim_{t \to 0^+} \frac{S(t)h - h}{t} \quad \text{exists in } H\}. \tag{1.4}$$

The infinitesimal generator A of the semigroup $S(t)$ is the operator

$$Ax = \lim_{t \to 0^+} \frac{S(t)x - x}{t}, \tag{1.5}$$

In the case of a uniformly continuous semigroup, we have the following characterization.

Theorem 1.1.6 (Theorem 1.1 in [28])**.** *A linear operator A is the infinitesimal generator of a uniformly continuous semigroup $S(t)$ on a Hilbert space if and only if $A \in L(H)$. Moreover, we have that*

$$S(t) = e^{tA} = \sum_{n=0}^{\infty} \frac{(tA)^n}{n!} \tag{1.6}$$

converges in the operator norm for every $t \geq 0$.

Example 1.1.7. *This decomposition is in line with Example 1.1.2 as for the generator to be the identity I, the Taylor expansion is exactly the one in Equation* (1.6)*. Similarly, even in the case of $\mathbb{R}^n$, for $n \in \mathbb{N}^+, n < \infty$, A being a linear operator, can be represented as a matrix. The matrix exponentiation agrees with the decomposition in* (1.6) *as well.*

We now state a few properties of the semigroup that will be useful.

Theorem 1.1.8 (Theorem 1.2 in [28])**.** *Let A be an infinitesimal generator of the C_0 semigroup $S(t)$ on a Hilbert space H. Then*

- *for $x \in H$,*

$$\lim_{h \to 0} \frac{1}{h} \int_t^{t+h} S(s)x ds = S(t)x. \tag{1.7}$$

- *for $x \in D(A), S(t)x \in D(A)$,*

$$\frac{d}{dt} S(t)x = AS(t)x = S(t)Ax. \tag{1.8}$$

- *for $x \in H, \int_0^t S(s)x ds \in D(a)$,*

$$A\left(\int_0^t S(s)x ds\right) = S(t)x - x. \tag{1.9}$$

1.2 Stochastic Calculus

1.2.1 Q-Wiener process

Given that the Brownian Motion is defined on the reals, we extend it to a separable Hilbert space, where it takes the name of Q-Wiener process. This ingredient is crucial for the proper extension of the standard Stochastic Differential Equations (SDEs) in the setting of Hilbert spaces.

Throughout this section we assume that all Hilbert spaces are infinite-dimensional, thereby enabling the analysis of stochastic partial differential equations whose governing linear operators are (possibly unbounded) differential operators.

First, we introduce some notation that will be adopted from now on; clarification will be given in case ambiguity might arise. Let $(H, \langle\cdot,\cdot\rangle)_H$ be a separable Hilbert space. Let $\{e_i\}_{i\in\mathbb{N}}$ an orthonormal basis of H. We denote by $\|\cdot\|_H$ the norm induced by the inner product.

H-valued Gaussian Random Variable

A crucial ingredient to properly define a Gaussian random variable is its covariance. For a finite-dimensional Gaussian random variable, its covariance is represented by the so-called covariance matrix K where $K_{i,j} = \mathbb{E}\left[(x_i - \mathbb{E}\left[x_i\right])(x_j - \mathbb{E}\left[x_j\right])\right]$, $1 \le j \le i \le n$ for a random vector $x = (x_1, \dots x_n)$. Naturally, we cannot extend this directly to the infinite case. In the same way that linear operators are the extension of matrices, we extend the notion of covariance to be a linear operator.

Definition 1.2.1. *Let $Q \in L(H)$, the trace of $[Q] := (QQ^\star)^{\frac{1}{2}}$ is defined as*

$$\mathrm{Tr}([Q]) = \sum_{i=0}^{\infty} \langle [Q]e_i, e_i\rangle_H. \tag{1.10}$$

Definition 1.2.2. *Denote $\mathcal{L}_1(H)$ the space of trace-class operator on H*

$$\mathcal{L}_1(H) := \left\{ Q \in L(H) : \mathrm{Tr}((QQ^\star)^{\frac{1}{2}}) < \infty \right\}. \tag{1.11}$$

The fact that the series converges from the assumption in Definition 1.2.2 is not proven here; see [16] for further details. We will show, however, that the trace value is independent of the basis choice, which is the content of the next proposition.

Proposition 1.2.3 (Lemma III.2 of [53])**.** *For an operator $Q \in L(H), \mathrm{Tr}([Q])$ is independent of the choice of the orthonormal basis $(e_i)_{i\in\mathbb{N}}$ of H.*

Proof. Qssume $(f_i)_{i\in\mathbb{N}}$ is another ONB (orthonormal basis) of H, Then by the sesquilinearity of the inner product, we can do the following.

$$\begin{aligned}\sum_{i,j}\langle [Q]e_i, f_j\rangle\langle f_j, e_i\rangle &= \sum_{i,j}\langle [Q]e_i, f_j\langle f_j, e_i\rangle\rangle \\ &= \sum_{i}\langle [Q]e_i, \sum_{j} f_j\langle f_j, e_i\rangle\rangle \\ &= \sum_{i}\langle [Q]e_i, e_i\rangle,\end{aligned}$$

where the exchange of the summation is justified by absolute convergence that follows directly from the definition of the trace class operator. By using the adjoint

$$\begin{aligned}\sum_{i,j}\langle [Q]e_i, f_j\rangle\langle f_j, e_i\rangle &= \sum_{i,j}\langle f_j, e_i\rangle\langle e_i, [Q]^\star f_j\rangle \\ &= \sum_{i}\langle f_j, [Q]^\star f_j\rangle = \sum_{j}\langle [Q]f_j, f_j\rangle\end{aligned}$$

and the statement of the proposition holds. □

Another important property of trace-class operators is compactness, which is a consequence of $[Q]$ having finite sum of eigenvalues. It is important to note that some authors set the compactness directly in the definition of the trace-class operators.

Corollary 1.2.4 (Proposition 4.6 of [22])**.** *If Q is a trace-class operator on H, then a is a compact operator.*

We now introduce the notion of a symmetric operator on a Hilbert space.

Definition 1.2.5. *Let Q be an operator acting on H, with a dense domain in H and domain $D(Q)$. Then we call Q symmetric if*

$$\langle Qx, y\rangle = \langle x, Qy\rangle, \quad \forall x, y \in D(Q). \tag{1.12}$$

This notion is closely related to that of a self-adjoint operator, that is when an operator $Q = Q^\star$, but there is a subtle distinction concerning the domain of the operator Q. Specifically, a symmetric operator satisfies $D(Q) \subset D(Q^\star)$, whereas for a self-adjoint operator, we require the stronger condition $D(Q) = D(Q^\star)$.

We now introduce the definition of a Q-Wiener process, which plays a central role in the construction of the stochastic integral in the Hilbert space setting.

Definition 1.2.6. *Let Q be a non-negative definite symmetric trace-class operator on H, let $\{e_i\}_{i\in\mathbb{N}}$ an orthonormal basis of H diagonalizing Q, i.e. Q can be represented as $Q(x) = \sum_{i=0}^{\infty}\lambda_i\langle x, e_i\rangle e_i$. In other words, the orthonormal basis $\{e_i\}_{i\in\mathbb{N}}$ forms the eigenvectors of Q and λ_i the corresponding eigenvalues. Let $\{w_i\}_{i=1}^{\infty}$ be a sequence of independent Brownian motions defined on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. The* *filtration* $\mathbb{F} := \{\mathcal{F}_t\}$ *satisfies the usual conditions (Appendix, Definition 5.2.2). The process* w_i *is adapted for every* $i \in \mathbb{N}$*. The process defined as follows is called Q-Wiener process.*

$$W_t := \sum_{j=1}^{\infty} \lambda_j^{\frac{1}{2}} w_j(t) e_j, \quad t \geq 0. \tag{1.13}$$

Note that $(W_t)_{t\geq 0}$ is a H-valued Stochastic process $\Omega \to H$. By slight abuse of notation, we denote the Q-Wiener process functional evaluation by

$$W_t(h) = \langle W_t, h \rangle = \sum_{j=1}^{\infty} \lambda_j^{\frac{1}{2}} w_j(t) \langle e_j, h \rangle. \tag{1.14}$$

W_t is a map from H to $L^2(\Omega, C([0,T], \mathbb{R}))$. We now give a brief result about the convergence of the series in Equation (1.13).

Lemma 1.2.7 (Corollary 2.3 in [28]). *In the settings of Definition 1.2.6, the series*

$$\sum_{j=1}^{\infty} \lambda_j^{\frac{1}{2}} w_j(t) e_j \tag{1.15}$$

converges $\mathbb{P}$ *a.s. uniformly.*

Proof. We will prove that the partial sums are Cauchy in $L^2(\Omega, C([0,T]), H)$. The convergence ensues from the completeness of the L^2 space. Define the partial sum $M_n := \sum_{j=1}^{n} \lambda_j^{\frac{1}{2}} w_j(t) e_j$. Note that $\|M_n - M_m\|$ is a submartingale for every $n \geq 0, m \leq n$. Therefore, by Doob inequality (5.5) with $p = 2$,

$$\mathbb{E}(\sup_{0\leq t\leq T} \|M_t^n - M_t^m\|_H^2) \leq 4\mathbb{E}\|M_t^n - M_t^m\|_H^2. \tag{1.16}$$

For the norm in $C([0,T], H)$ we simply take the sup norm, therefore $\|W_t\|_{L^2(\Omega, C([0,T],H))} = \mathbb{E}\left[\sup_{0\leq t\leq T} \|W_t\|_H^2\right]$. We can now bound the norm of the partial sums.

$$\begin{aligned}
\mathbb{E}(\sup_{0\leq t\leq T} \|M_t^n - M_t^m\|_H^2) &\leq 4\mathbb{E}\|M_t^n - M_t^m\|_H^2 \\
&\leq 4\mathbb{E}\langle \sum_{i=m}^{n} \lambda_i^{\frac{1}{2}} w_i(t) e_i, \sum_{j=m}^{n} \lambda_j^{\frac{1}{2}} w_j(t) e_j \rangle \\
&\leq 4 \sum_{i=m}^{n} |\lambda_i| \mathbb{E}\left[w_i(t)^2\right] \\
&= 4t \sum_{i=m}^{n} |\lambda_i|.
\end{aligned}$$

Since Q is a trace-class operator, we have that $\lim_{n,m\to\infty} 4t\sum_{i=m}^{n}|\lambda_i| \to 0$ implying that the partial sums M_t^n are Cauchy in $L^2(\Omega, C([0,T],H))$. This implies directly from the completeness of the Hilbert space that the whole sum converges in L^2, therefore, we have that

$$\left\|\sum_i^\infty \lambda_i^{\frac{1}{2}} w_i(t)e_i\right\| < \infty, \quad 0 \le t \le T. \tag{1.17}$$

Using Hölder inequality, we have that

$$\|W_t^n - W_t^m\|_{L^1} \le \mu(\Omega)^{\frac{1}{2}}\|W_t^n - W_t^m\|_{L^2} \to 0. \tag{1.18}$$

Therefore, we have converge in L^1 sense, moreover, since μ is a probability measure it is obviously finite; all the assumptions of the Vitali Theorem are in place (Appendix, Theorem 5.2.8). Using the Vitali convergence Theorem, we get that $\sup_{0\le t\le T}\|M_t^n - M_t^m\|_H^2 \to 0$ implies convergence in probability. □

We now state a theorem that shows that the properties of the standard Wiener process extends intuitively to the Q-Wiener process.

Theorem 1.2.8 (Theorem 2.1 in [28]). *An H-valued Q-Wiener process W_t, $t \ge 0$ has the following properties.*

- $W_0 = 0$
- *W_t has continuous trajectories a.s.*
- *W_t has independent increments.*
- *The covariance for any $h, h' \in H$ and $s, t \ge 0$ of W_t is given by*

$$\mathbb{E}\left[W_t(h)W_s(h')\right] = (s \wedge t)\langle Qh, h'\rangle. \tag{1.19}$$

- *For any $h \in H$, the increments are normally distributed, i.e.*

$$(W_t - W_s)(h) \sim N(0, (t-s)\langle Qh, h\rangle). \tag{1.20}$$

The proof follows directly from the Definition of $(W_t)_{t\ge 0}$ and the properties of the independent Brownian Motion w_j.

1.2.2 Hilbert-space-valued martingale

Before introducing the stochastic Differential Equations on H, we first introduce the notion of a martingale in the Hilbert space setting. This concept is crucial to be able to properly define the stochastic integral for a large class of stochastic processes, bigger than just elementary processes.

Definition 1.2.9. *For $T > 0$, consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with its associated filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t \le T}$. Consider the H-valued process $(M_t)_{0 \le t \le t}$, adapted to the filtration $\mathbb{F}$. Furthermore, we assume that $(M_t)_{t \le T}$ is integrable, i.e. $\mathbb{E}\|M_t\|_H < \infty$. $(M_t)_{t \le T}$ is called a martingale if for any $0 \le s \le t$,*

$$\mathbb{E}\left[M_t | \mathcal{F}_s\right] = M_s \quad \mathbb{P}\text{-a.s.} \tag{1.21}$$

where the conditional expectation is understood in the Bochner sense.

Remark 1.2.10. *Note that the functional $x \mapsto \langle x, h \rangle_H$ is continuous for every $h \in H$, and thus Borel measurable by the definition of continuity through open sets. Consequently, $\langle M_t(\cdot), h \rangle_H$ is Borel measurable provided $(M_t)_{t \le T}$ itself is Borel measurable.*

With this in mind, we can see that, given the assumptions of Definition 1.2.9, we have that $\langle M_t, h \rangle$ is measurable with respect to $\mathcal{F}_t / B(\mathbb{R})$ for all $h \in H$. This implies directly the measurability of $\|M_t\|$ for every $t \le T$. The converse is also true, but we will omit the details here. Both these claims allow us to define an equivalent definition for M_t to be a martingale, i.e.

$$\mathbb{E}\left[\langle M_t, h \rangle | \mathcal{F}_s\right] = \langle M_s, h \rangle \quad \forall h \in H \quad \mathbb{P}\text{-a.s.} \tag{1.22}$$

Furthermore, if the process $(M_t)_{t \ge 0}$ is a H-valued martingale, the process $(\|M_t\|_H)_{t \le T}$ is a real valued submartingale. This indeed follows from the Jensen's inequality applied to the convex function $\|\cdot\|_H$.

1.2.3 Stochastic integral with respect to Q-Wiener process

We are now interested in the definition and the properties of the Itô integral in the setting of Hilbert space-valued processes and, in particular, the Q-Wiener process that we defined in Section 1.2.1.

Stochastic Itô integral for elementary processes

As standard procedure, we begin by defining the integral for elementary (or simple) processes, then we extend it to a larger class of Stochastic processes using the Itô Isometry.

We start by introducing K, H two separable Hilbert spaces, and Q be a symmetric nonnegative definite trace-class operator on K. Without loss of generality, we assume that all the eigenvalues of Q are positive. In fact, if that is not the case, we consider $K' = \ker Q^{\perp}$ as of K. Clearly, this subspace satisfies such condition and Q remains well-defined as an operator on K'. We denote by $\{e_i\}_{i \in \mathbb{N}}$ the corresponding eigenvectors that form an ONB of K.

Proposition 1.2.11 (2.2 of [28]). *The space $K_Q := Q^{\frac{1}{2}} K$ equipped with the scalar product*

$$\langle u, v \rangle_{K_Q} = \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle u, e_i \rangle_K \langle v, e_i \rangle_K \tag{1.23}$$

is a separable Hilbert space with ONB being $\{\lambda_i^{\frac{1}{2}} e_i\}_{i \in \mathbb{N}}$ where $\{e_i\}_{i \in \mathbb{N}}$ is the set of eigenvectors of Q.

Furthermore, note that Q is compact, see Proposition 1.2.4. Moreover, since Q is self-adjoint, it is also normal (i.e. $QQ^\star = Q^\star Q$), then, due to spectral Theorem IX.2.2 of [22], we have that $Q^{\frac{1}{2}}$ is well defined.

Proof. The fact that the basis is dense comes directly from the fact that $\{e_i\}_{i\in\mathbb{N}}$ is an ONB in K, and their image is $\{\lambda_i^{\frac{1}{2}} e_i\}_{i\in\mathbb{N}}$ by the spectral mapping Theorem (Theorem 4.10 of [22]) for compact operators, since Q is compact. In fact, $\{e_i\}_{i\in\mathbb{N}}$ are obviously pairwise disjoint and $\langle \lambda_i^{\frac{1}{2}} e_i, \lambda_i^{\frac{1}{2}} e_i\rangle = \frac{1}{\lambda_i}|\lambda_i|\|e_i\| = 1$. The completeness of K_Q comes directly from the fact that K is complete and $Q^{\frac{1}{2}}$ is linear. In fact, take u_n^Q Cauchy in K_Q, then $\|u_n^Q - u_m^Q\| \to 0$ as $n, m \to \infty$. Then, $\|u_n^Q - u_m^Q\|_{K^Q} = \|Q^{\frac{1}{2}} u_n - Q\frac{1}{2} u_m\|_{K^Q}$ for $u_n, u_m \in K$. We can conclude since $\|Q^{\frac{1}{2}} u_n - Q\frac{1}{2} u_m\|_K \leq \|Q^{\frac{1}{2}}\|_{L(K)} \|u_n - u_m\|_K \to 0$ and noting that $\|Qu\|_{K_Q} = \|Qu\|_K$. □

We now introduce an important function space. Let H_1, H_2 two real separable Hilbert spaces and $\{e_i\}_{i\in\mathbb{N}}$ ONB of H_1, the Hilbert-Schmidt operators from H_1 to H_2 are defined as

$$\mathcal{L}_2(H_1, H_2) := \left\{ L \in L(H_1, H_2) : \sum_{i=1}^{\infty} \|Le_i\|_{H_2}^2 < \infty \right\}. \tag{1.24}$$

Together with the norm

$$\|L\|_{\mathcal{L}_2(H_1,H_2)} = \left(\sum_{i=1}^{\infty} \|Le_i\|_{H_2}^2 \right)^{1/2}, \tag{1.25}$$

the $\mathcal{L}_2(H_1, H_2)$ is a Hilbert space, see [54] for the proof. For K_Q being H_1 and H being H_2, applying the definition of norm, we see that

$$\|L\|_{\mathcal{L}_2(K_Q,H)} = \operatorname{Tr}\left((LQ^{1/2})^\star (LQ^{1/2}) \right)^{\frac{1}{2}}. \tag{1.26}$$

Congruently with the definition of norm induced by the inner product, we defined the inner product to be

$$\langle L, M\rangle_{\mathcal{L}_2(K_Q,H)} = \operatorname{Tr}\left((LQ^{1/2})(MQ^{1/2})^\star \right). \tag{1.27}$$

We now introduce the definition of elementary processes. In the same manner as the standard finite-dimensional Itô integral, we begin by defining the integration for a simple (or elementary) process, whose definition is equivalent to the real case.

Definition 1.2.12. *For two separable Hilbert spaces H, K, we denote by $\mathcal{E}(\mathcal{L}(K, H))$ the class of elementary processes adapted to the filtration $\mathbb{F}$ of the form*

$$\phi(t, \omega) = \phi_0(\omega)\mathbb{1}_0(t) + \sum_{j=0}^{n-1} \phi_j(\omega), \mathbb{1}_{(t_j,t_{j+1}]} \tag{1.28}$$

where we defined the partition $\pi := 0 \leq t_0 < \cdots < t_n = T$. The functions $\phi_i : \Omega \to \mathcal{L}_2(K_Q, H)$ are $\mathcal{F}_{t_i}$ measurable functions for $0 \leq i \leq n-1$. Moreover, ϕ is bounded if it is bounded in $\mathcal{L}_2(K_Q, H)$.

With the definition of elementary processes in place, we are now ready to define the Itô integral for such processes intuitively.

Definition 1.2.13. *For an elementary process $\phi \in \mathcal{E}(\mathcal{L}(K,H))$, we define the stochastic Itô integral,* $\mathrm{Int}(\phi)$ *which is an H-valued stochastic process as follows:*

$$\int_0^t \phi(s) dW_s := \sum_{j=0}^{n-1} \phi_j (W_{t_{j+1}\wedge t} - W_{t_j \wedge t}) \tag{1.29}$$

The first result is about the Itô isometry for such processes.

Proposition 1.2.14 (Theorem 2.3 in [29]). *For bounded elementary processes $\phi \in \mathcal{E}(L(K,h))$ and for $0 \leq t \leq T$*

$$\mathbb{E}\left\| \int_0^t \Phi(s)\, dW_s \right\|_H^2 = \mathbb{E} \int_0^t \|\Phi(s)\|^2_{\mathcal{L}_2(K_Q,H)}\, ds < \infty. \tag{1.30}$$

The parallelism with the real finite case continues. The objective now is to show that elementary processes approximate a much richer class of processes and use the Itô isometry (Proposition 1.2.14) to extend the integral definition to such a class to be the isometric linear extension.

We now define the class of adapted, measurable processes satisfying the condition

$$\mathbb{E} \int_0^t \|\Phi(s)\|^2_{\mathcal{L}_2(K_Q,H)}\, ds < \infty. \tag{1.31}$$

This will enable us to properly extend the integral while maintaining its well-defined nature.

Definition 1.2.15. *Let $\Lambda_2(K_Q,H)$ be the set of $\mathcal{L}_2(K_Q,H)$ processes, such that are measurable maps from $([0,T] \times \Omega, \mathcal{B}([0,T]) \otimes \mathcal{F}) \to (\mathcal{L}_2(K_Q,H), \mathcal{B}(\mathcal{L}_2(K_Q,H)))$ and adapted to the filtration $\mathbb{F} := \{\mathcal{F}_t\}_{t\leq T}$ and satisfying the following condition*

$$E \int_0^T \|\Phi(t)\|^2_{\mathcal{L}_2(K_Q H)}\, dt < \infty. \tag{1.32}$$

We define the corresponding norm

$$\|\Phi\|_{\Lambda_2(K_Q,H)} = \left(E \int_0^T \|\Phi(t)\|^2_{\mathcal{L}_2(K_Q,H)}\, dt \right)^{1/2}. \tag{1.33}$$

Equipped with such norm, the space $\Lambda_2(K_Q,H)$ is a Hilbert space. We omit the proof here, but one can prove the completeness of the space from the completeness of the two underlying spaces, K and H.

The next proposition shows that the elementary processes $\mathcal{E}(\mathcal{L}(K,H))$ are dense in $\Lambda_2(K_Q,H)$.

Proposition 1.2.16 (Proposition 2.2 in [28]). *For $\Phi \in \Lambda_2(K_Q, H)$, there exists a sequence of bounded elementary processes $\Phi_n \in \mathcal{E}(\mathcal{L}(K,H))$ that approximate Φ,i.e.*

$$\|\Phi_n - \Phi\|^2_{\Lambda_2(K_Q,H)} = E\int_0^T \|\Phi_n(t) - \Phi(t)\|^2_{\mathcal{L}_2(K_Q,H)}\, dt \xrightarrow{n\to\infty} 0 \tag{1.34}$$

The proof is technical and lengthy; therefore, for the details consult the reference. This is a crucial result that will allow us, together with the fact that the Itô integral w.r.t elementary process is an isometry, to properly define the integral for $\Lambda_2(K_Q, H)$, which contains the typical process we aim at working with. In the next section, we formalize how to properly extend the Itô integral to such a class.

Stochastic Itô integral

Definition 1.2.17. *For a process $\Phi \in \Lambda_2(K_Q, H)$, the stochastic integral with respect to a K-valued Q-Wiener process W_t is the unique isometric linear extension of the mapping*

$$\Phi(\cdot) \to \int_0^T \Phi(s)dW_s. \tag{1.35}$$

In other words, for a process $\Phi \in \Lambda_2(K_Q, H)$ take an approximating sequence of elementary processes Φ_n such that $\|\Phi - \Phi_n\|_{\Lambda_2(K_Q,H)} \to 0$ as $n \to \infty$. Then the integral of Φ corresponds to

$$\mathrm{Int}(\Phi) = \lim_{n\to\infty}\int_0^T \Phi_n dW_s. \tag{1.36}$$

By the isometry and the fact that it maps to $L^2(\Omega, H)$, the limit exists and it is unique. For the integration of $\Phi \in \Lambda_2(K_Q, H)$ from 0 to $t < T$, we define the integral using the indicator function as such

$$\int_0^t \Phi(s)dW_s := \int_0^T \Phi(s)\mathbb{1}_{[0,t]}dW_s. \tag{1.37}$$

We now have a proposition extending Proposition 1.2.14 showing that the Itô integral with respect to a Q- Wiener process is a martingale.

Proposition 1.2.18 (Theorem 2.3 in [28]). *For $\Phi \in \Lambda_2(K_Q, H)$, the stochastic integral with respect to a K-valued Q-Wiener process W_t is an isometry between $\Lambda_2(K_Q, H)$ and the space $\mathcal{M}^2_T(H)$ denoting the continuous squared integrable martingales on H, i.e.*

$$E\left\|\int_0^t \Phi(s)\, dW_s\right\|^2_H = E\int_0^t \|\Phi(s)\|^2_{\mathcal{L}_2(K_Q,H)}\, ds < \infty \tag{1.38}$$

Proof. Note that for an elementary process, the Itô integral is square integrable. The martingale part follows directly from the measurability of ϕ_j and W_{t_j}, in fact, for $t_m =$

$\max_n t_n \leq s$ and by the tower property

$$\begin{aligned}
\mathbb{E}\left[\sum_{j=0}^{n-1} \phi_j (W_{t_{j+1}} - W_{t_j}) | \mathcal{F}_s\right] &= \sum_{j=0}^{m} \phi_j (W_{t_{j+1}} - W_{t_j}) + \mathbb{E}\left[\sum_{j=m+1}^{n-1} \phi_j (W_{t_{j+1}} - W_{t_j}) | \mathcal{F}_s\right] \\
&= \sum_{j=0}^{m} \phi_j (W_{t_{j+1}} - W_{t_j}) + \\
&+ \sum_{j=m+1}^{n-1} \mathbb{E}\left[\mathbb{E}\left[\phi_j (W_{t_{j+1}} - W_{t_j}) | \mathcal{F}_{t_j}\right] | \mathcal{F}_s\right] \\
&= \sum_{j=0}^{m} \phi_j (W_{t_{j+1}} - W_{t_j}).
\end{aligned}$$

We now prove Equation (1.38). Take an approximating sequence of elementary processes Φ_n, then by a triangular argument, we can assume

$$\|\Phi_{n+1} - \Phi_n\|_{\Lambda_2(K_Q,H)} \leq \frac{1}{2^n}. \tag{1.39}$$

Using Doob inequality 5.5 we get

$$\begin{aligned}
\sum_{n=1}^{\infty} P\left(\sup_{t\leq T} \left\| \int_0^t \Phi_{n+1}(s)\, dW_s - \int_0^t \Phi_n(s)\, dW_s \right\|_H > \frac{1}{n^2}\right) \\
\leq \sum_{n=1}^{\infty} n^4 E \left\| \int_0^T \left(\Phi_{n+1}(s) - \Phi_n(s)\right) dW_s \right\|_H^2 \\
= \sum_{n=1}^{\infty} n^4 E \int_0^T \|\Phi_{n+1}(s) - \Phi_n(s)\|^2_{\mathcal{L}_2(K_Q,H)}\, ds \leq \sum_{n=1}^{\infty} \frac{n^4}{2^n}.
\end{aligned}$$

Since $\sum_{n=1}^{\infty} \frac{n^4}{2^n} < \infty$, by applying the Borel-Cantelli Lemma, we have that

$$P\left(\lim_{n\to\infty} \sup_{t\leq T} \left\| \int_0^t \Phi_{n+1}(s)\, dW_s - \int_0^t \Phi_n(s)\, dW_s \right\|_H > \frac{1}{n^2}\right) = 0. \tag{1.40}$$

Thus,

$$\sum_{n=1}^{\infty} \left(\int_0^t \Phi_{n+1}(s)\, dW_s - \int_0^t \Phi_n(s)\, dW_s \right) \tag{1.41}$$

is Cauchy and converges to $\int_0^t \Phi(s) dW_s$ in $L^2(\Omega, H)$ for every $t \leq T$.

□

Moreover, as a corollary, we have that

$$P\left(\sup_{t\leq T} \left\| \int_0^t \Phi_n(s)\, dW_s - \int_0^t \Phi(s)\, dW_s \right\|_H \to 0\right) = 1. \tag{1.42}$$

We successfully defined the Stochastic Itô Integral for integrands $\Phi \in \Lambda_2(K_Q, H)$. However, this is a quite restrictive class of Stochastic processes. Thus, we will extend the definitsion to the following class of processes.

Definition 1.2.19. *Let $\mathcal{P}(K_Q, H)$ be the set of $\mathcal{L}_2(K_Q, H)$-valued processes $\Phi = (\Phi_t)_{t\in[0,T]}$ that are $\mathbb{F}$-adapted and satisfy*

$$\Phi : \big([0,T] \times \Omega,\ \mathcal{B}([0,T]) \otimes \mathcal{F}_T\big) \longrightarrow \big(\mathcal{L}_2(K_Q, H),\ \mathcal{B}(\mathcal{L}_2(K_Q, H))\big)$$

is measurable, and

$$\mathbb{E}\int_0^T \|\Phi_t\|^2_{\mathcal{L}_2(K_Q,H)}\, \mathrm{d}t < \infty.$$

It is immediately clear from the definition, together with the Markov inequality, that $\Lambda_2(K_Q, H)$ is a subset of $\mathcal{P}(K_Q, H)$.

To formalize our goal, for an approximating sequence $\Phi_n \in \mathcal{E}(L(K, H))$ such that $P\left(\int_0^T \|\Phi_n(t,\omega) - \Phi(t,\omega)\|^2_{\mathscr{L}_2(K_Q,H)}\, dt > 0\right) \to 0$, we want to show that $\int_0^T \Phi_n(t)\, dW_t \to \int_0^T \Phi(t)\, dW_t$. The procedure is similar to the above, approximate an arbitrary process in $\mathcal{P}(K_Q, H)$ by a sequence in $\Lambda_2(K_Q, H)$, the integral of which we properly defined above. To do so, we need a few asymptotic results.

Lemma 1.2.20 (Lemma 2.3 in [28])**.** *Let $\Phi \in \mathcal{P}(K_Q, H)$, then there exists an approximating sequence of bounded processes $\Phi_n \in \mathcal{E}(K_Q, H)$ such that*

$$\int_0^T \|\Phi(t,\omega) - \Phi_n(t,\omega)\|^2_{\mathcal{L}_2(K_Q,H)}\, dt \to 0 \quad P - a.s. \tag{1.43}$$

By applying the same argument as in the proof of the previous Lemma, we obtain the following result:

Lemma 1.2.21 (Lemma 2.4 in [28])**.** *Let $\Psi(t), t \leq T$ be an H-valued, $\mathcal{F}_t$-adapted stochastic process such that*

$$P\left(\int_0^T \|\Psi(t)\|_H\, dt < \infty\right) = 1. \tag{1.44}$$

Then there exists a sequence of bounded elementary processes $\Phi_n \in \mathcal{E}(H)$ such that

$$\int_0^t \|\Phi(t,\omega) - \Phi_n(t,\omega)\|_H dt \to 0 \quad P - a.s. \tag{1.45}$$

The final Lemma that we need is the following, which we state without proof, as it is not insightful for our goal.

Lemma 1.2.22 (Lemma 2.5 in [28])**.** *Let $\Phi \in \Lambda_2(K_Q, H)$, then for an arbitrary $\delta > 0$ and $n > 0$,the following hold true*

$$P\left(\sup_{t\leq T}\left\|\int_0^t \Phi(s)\, dW_s\right\|_H > \delta\right) \leq \frac{n}{\delta^2} + P\left(\int_0^T \|\Phi(s)\|^2_{\Lambda_2(K_Q,H)}\, ds > n\right). \tag{1.46}$$

We are finally ready for the main result of this section, the construction of the Stochastic integral in $\mathcal{P}(K_Q, H)$.

Lemma 1.2.23 (Lemma 2.5 in [29])**.** *Let Φ_n be a sequence in $\Lambda_2(K_Q, H)$ such that*

$$P\left(\int_0^T \|\Phi_n(t,\omega) - \Phi(t,\omega)\|^2_{\mathcal{L}_2(K_Q,H)}\, dt > 0\right) \to 0 \tag{1.47}$$

for $\Phi \in \mathcal{P}(K_Q, H)$. Then, there exists an H-valued $\mathcal{F}_T$ measurable random variable called $\int_0^T \Phi(t) dW_t$ such that

$$\int_0^T \Phi_n(t) dW_t \to \int_0^T \Phi(t) dW_t \tag{1.48}$$

in probability. Moreover, $\int_0^T \Phi(t) dW_t$ does not depend on the choice of the approximating sequence Φ_n. The random variable $\int_0^T \Phi(t) dW_t$ is then called the stochastic integral of $\Phi(t)$ with respect to the Q-Wiener process W_t.

Proof. By the assumption, for any $\epsilon > 0$, it holds true that

$$\lim_{n,m\to\infty} P\left(\int_0^T \|\Phi_n(t,\omega) - \Phi_m(t,\omega)\|^2_{\mathcal{L}_2(K_Q,H)} > \epsilon\right) = 0. \tag{1.49}$$

By applying Lemma 1.2.22 to $\delta > 0$

$$\begin{aligned}
\limsup_{m,n\to\infty} &P\left(\left\|\int_0^T \Phi_n(t)\, dW_t - \int_0^T \Phi_m(t)\, dW_t\right\|_H > \delta\right) \\
&\leq \frac{\epsilon}{\delta^2} + \lim_{n,m\to\infty} P\left(\int_0^T \|\Phi_n(t,\omega) - \Phi_m(t,\omega)\|^2_{\mathcal{L}_2(K_Q,H)} > 0\right) = \frac{\epsilon}{\delta^2}
\end{aligned}$$

Since it holds for every $\epsilon > 0$, then we conclude that

$$\lim_{m,n\to\infty} P\left(\left\|\int_0^T \Phi_n(t)\, dW_t - \int_0^T \Phi_m(t)\, dW_t\right\|_H > \delta\right) = 0. \tag{1.50}$$

Thus, the conclusion follows immediately. $\square$

1.2.4 Itô's Lemma, Martingale Representation Theorem and stochastic Fubini

Itô's Lemma

The content of the following Theorem requires knowledge of calculus in Hilbert space, such as Bochner Integration and Fréchet derivative, which are discussed in Section 5.4 of the Appendix.

For a stochastic process $X(t)$, we are interested in the dynamics of $F(t, X(t))$ for a differentiable function F. This is precisely the content of Itô's Lemma.

Theorem 1.2.24 (Theorem 2.9 in [28]). *Let Q be a symmetric nonnegative trace-class operator on a separable Hilbert space K and let $\{W_t\}_{t\leq T}$ be a Q-Wiener process on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{F} = \{\mathcal{F}_t\}_{t\leq T}$. Let $\Psi(s)$ be an H valued random variable where H is a separable Hilbert space, and $\Psi(s)$ is $\mathcal{F}_s$-measurable, $\mathbb{P}$ -a.s. Bochner integrable on the interval $[0,T]$. Finally, let $\Phi \in \mathcal{P}(K_Q, H)$. Let $F : [0,T] \times H \to \mathbb{R}$ be a continuous function with continuous and bounded Fréchet partial derivatives F_t, F_x, F_{xx}. Then, for a stochastic process of the form*

$$X(t) = X(0) + \int_0^t \Psi(s)\, ds + \int_0^t \Phi(s)\, dW_s, \tag{1.51}$$

the so-called ItôFormula holds true

$$\begin{aligned} F(t, X(t)) = F(0, X(0)) &+ \int_0^t \langle F_x(s, X(s)), \Phi(s)\, dW_s\rangle_H \\ &+ \int_0^t F_t(s, X(s)) + \langle F_x(s, X(s)), \Psi(s)\rangle_H ds \\ &+ \frac{1}{2}\int_0^t \operatorname{tr}\left[F_{xx}(s, X(s))(\Phi(s)Q^{1/2})(\Phi(s)Q^{1/2})^\star\right] ds \end{aligned}$$

P-a.s. for all $0 \leq t \leq T$.

For the proof, consult Theorem 2.9 in [29].

Martingale Representation Theorem

This section aims to prove the martingale representation Theorem for Hilbert-space valued processes. For the standard version of the martingale representation Theorem (in the more general settings of local martingales) check Section 8.1 of [55]. This Theorem is crucial for the existence of a solution in BSDE, and it is heavily used in other proofs as well, which we will see later in this work.

Denote by $\mathcal{F}_t^j$ the sigma algebra generated by the j-th Wiener process, i.e.

$$\mathcal{F}_t^j \coloneqq \sigma\{w_j(s) : s \leq t\}. \tag{1.52}$$

Moreover, define

$$\mathcal{G}_t \coloneqq \sigma\left\{\bigcup_{j=1}^{\infty} \mathcal{F}_t^j\right\}. \tag{1.53}$$

Similarly, we define the version for W_t by

$$\mathcal{F}_t^W = \sigma\left(\{W_s(k) : k \in K, \text{ and } s \leq t\}\right). \tag{1.54}$$

Proposition 1.2.25. *If $\mathcal{F}_i$ are independent σ-algebra, then*

$$L^2\left(\Omega, \sigma\left\{\bigcup_{i=1}^{\infty} \mathcal{F}_i\right\}, P\right) = \bigoplus_{i=1}^{\infty} L^2(\Omega, \mathcal{F}_i, P). \tag{1.55}$$

Proof. We prove only the $\subset$ direction since the other inclusion is straightforward. Assume $f \in L^2\left(\Omega, \sigma\left\{\bigcup_{i=1}^{\infty} \mathcal{F}_i\right\}, P\right)$, then $f \in L^2(\Omega, \mathcal{F}_i, P)$ for ever $i \in \mathbb{N}$. Define $f_i :=\mathbb{E}\left[f|\mathcal{F}_i\right]$ and let $F := \sum_{i=0}^{\infty} f_i$. By the independence of the σ-fields $\mathcal{F}_i$, we get that $F = f$, thus $f \in \bigoplus_{i=1}^{\infty} L^2(\Omega, \mathcal{F}_i, P)$. This concludes the proof. $\square$

Before stating the main theorem of this section, we present an intermediate lemma that will be crucial in the proof of the martingale representation Theorem.

Lemma 1.2.26 (Theorem 4.3.3 in [49])**.** *Every real-valued $\mathcal{F}_t^W$ martingale m_t in $L^2(\Omega, \mathcal{F}_t^w, P)$ has the unique (up to modification) representation*

$$m_t(\omega) = \mathbb{E}m_0 + \sum_{j=1}^{\infty} \int_0^t \lambda_j^{1/2}\, \phi_j(s,\omega)\, dw_j(s), \tag{1.56}$$

and $\sum_{j=1}^{\infty} \lambda_j E \int_0^T \phi_j^2(s,\omega)\, ds < \infty$.

We now come to the main result of this section, the martingale representation Theorem I.

Theorem 1.2.27 (Theorem 2.5 in [28])**.** *Let H, K be two separable Hilbert spaces, W_t be the K-valued Q-Wiener process and M_t an H-valued continuous squared-integrable martingale with respect to the filtration $\mathcal{F}_t^W$. Then there exists a process $\Phi \in \Lambda_2(K_Q, H)$ such that*

$$M_t = \mathbb{E}\left[M_0\right] + \int_0^t \Phi(s) dW_s. \tag{1.57}$$

Proof. Let $\{e_i\}_{i\in\mathbb{N}}$ be the ONB of H and $\{f_i\}_{i\in\mathbb{N}}$ the ONB of K. In the view of the Lemma 1.2.26, there exists a process $\phi_j^i(t,\omega)$ such that

$$\langle M_t, e_i\rangle = \mathbb{E}\left[\langle M_0, e_i\rangle_H\right] + \sum_{j=1}^{\infty} \int_0^t \lambda_j^{\frac{1}{2}} \phi_j^i(s,\omega) dw_j(s). \tag{1.58}$$

Note that what we did is formally correct as, due to (1.22), the inner product of a martingale is a real-value martingale, thus we can apply the Lemma 1.2.26. Moreover, by Theorem 4.13.d of [22], we have $M_t = \sum_{i=1}^{\infty} \langle M_t, e_i\rangle_H e_i$. Inserting Equation (1.58) in the above, we get

$$M_t = \mathbb{E}\sum_{i=1}^{\infty} \langle M_0, e_i\rangle_H e_i + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \int_0^t \lambda_j^{1/2}\, \phi_j^i(s,\omega)\, e_i\, dw_j(s). \tag{1.59}$$

Using the fact that M_t is square integrable, together with the standard Fubini Theorem, we can interchange the summations in the above. Together with the fact that $\mathbb{E}\sum_{i=1}^{\infty} \langle M_0, e_i\rangle_H e_i = \mathbb{E}\left[M_0\right]$, we get

$$M_t = \mathbb{E}M_0 + \sum_{j=1}^{\infty} \lambda_j^{1/2} \sum_{i=1}^{\infty} \int_0^t \phi_j^i(s,\omega)\, e_i\, dw_j(s). \tag{1.60}$$

For $h \in H$ and $k \in K_Q$, we define $\Phi(s,\omega)$ such that

$$\langle \Phi(s,\omega)k, h\rangle_H = \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} \lambda_j \langle h, e_i\rangle_H \langle k, f_j\rangle_{K_Q} \phi_j^i(s,\omega). \tag{1.61}$$

Given that the right hand side is measurable and L^2 integrable, we have that $\Phi(s,\omega) \in \Lambda_2(K_Q, H)$. The final step is to replace the right-hand side of (1.60) by $\int_0^t \Phi(s,\omega)dW_s$ by the definition of the stochastic integral. □

Stochastic Fubini

Another significant result is the stochastic counterpart of the Fubini Theorem, which lets us interchange a stochastic integral with a deterministic one. Here we only present the result and some intuition for the proof, but we refer to Theorem 2.8 of [29] for a detailed proof.

Theorem 1.2.28 (Theorem 2.8 in [29]). *Let $(\Omega, \mathcal{F}, \mathbb{F} := \{\mathcal{F}\}_{t\le T}\mathbb{P})$ a probability space and let $(G, \mathcal{G}, \mu)$ be a finite measurable space ($\mu(G) < \infty$) and $\Phi : ([0,T]\times\Omega\times G, B([0,T])\otimes\mathcal{F}_T\otimes\mathcal{G}) \to (H, \mathcal{B}(H))$ measurable with respect to $(\mathcal{B}([0,T]), \mathcal{F}_T, \mathcal{G})$ such that for every $x \in G$, the stochastic process $\Phi(\cdot,\cdot,x)$ is adapted to the filtration $\mathbb{F}$. Moreover, we let W_t be a Q- Wiener process on the probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\le T}, \mathbb{P})$. Therefore, if*

$$|||\Phi||| := \int_G \|\Phi(\cdot,\cdot,x)\|_{\Lambda_2(K_Q,H)}\, \mu(dx) < \infty, \tag{1.62}$$

then the following holds true:

- *$\int_0^T \Phi(t,\cdot,x)dW_t$ has a measurable version viewed as a map $(\Omega\times G, \mathcal{F}_T\otimes\mathcal{G}) \to (H, \mathcal{B}(H))$.*
- *$\int_G \Phi(\cdot,\cdot,x)\mu(dx)$ is $\{\mathcal{F}_t\}_{t\le T}$ adapted.*
- *The following holds true $\mathbb{P}$ a.s.*

$$\int_G \left(\int_0^T \Phi(t,\cdot,x)\, dW_t\right)\mu(dx) = \int_0^T \left(\int_G \Phi(t,\cdot,x)\mu(dx)\right) dW_t. \tag{1.63}$$

Proof. We only sketch a proof, collecting the main ingredients needed to prove the third equality. Consider an approximating bounded sequence Φ_n

$$\Phi_n(t,\omega,x) := \begin{cases} \dfrac{n\,\Phi(t,\omega,x)}{\|\Phi(t,\omega,x)\|_{\mathcal{L}_2(K_Q,H)}} & \text{if } \|\Phi(t,\omega,x)\|_{\mathcal{L}_2(K_Q,H)} > n, \\ \Phi(t,\omega,x) & \text{otherwise.} \end{cases} \tag{1.64}$$

By the Dominate Converge Theorem, we have that $\Phi_n \to \Phi$ in $\Lambda_2(K_Q,H)$ norm. By the Itô isometry, it follows that

$$\lim_{n\to\infty}\int_0^T \Phi_n(t,\cdot,x)dW_t = \int_0^T \Phi(t,\cdot,x)dW_t. \tag{1.65}$$

But this directly implies the convergence in $|||\cdot|||$ sense. We now prove Equation (1.63) by proving that the following convergence results

$$E\left\|\int_G\left(\int_0^T \Phi_n(t,\cdot,x)\,dW_t\right)\mu(dx)-\int_G\left(\int_0^T \Phi(t,\cdot,x)\,dW_t\right)\mu(dx)\right\|_H \tag{1.66}$$

and

$$E\left\|\int_0^T\left(\int_G \Phi_n(t,\cdot,x)\,\mu(dx)\right)dW_t-\int_0^T\left(\int_G \Phi(t,\cdot,x)\,\mu(dx)\right)dW_t\right\|_H. \tag{1.67}$$

Equation (1.66) converges to zero directly from the fact that $|||\Phi-\Phi_n|||\to 0$. With a bit more work, Equation (1.67) follows as well from the convergence in the $|||\cdot|||$ sense.

□

1.3 Stochastic Partial Differential Equation

In this chapter, we apply the theory developed in the previous chapters to define Stochastic Partial Differential Equations (SPDEs) and explore various concepts of solutions, with a focus on the case of mild solutions. SPDEs ensue directly upon formally defining SDE in infinite dimensions; in the infinite-dimensional context, the derivative and integral are linear operators. SPDE is a highly active research topic, and due to the vast amount of material that can be covered on the subject of SPDE, we attempt to limit the results presented to those relevant to the scope of this work.

Furthermore, we will explore how SPDEs are utilized to model the prices of underlying assets, and in particular, their application in the context of forward pricing.

We begin by defining the general form of a Stochastic Differential Equations, then present the general assumptions under which we will consider two types of solutions, namely the strong and the mild solutions.

Throughout this chapter, similarly to previous chapters, we consider H, K to be two separable Hilbert spaces, $(W_t)_{t\geq 0}$ a K-values Q-Wiener process on the probability space $(\Omega,\mathcal{F},\mathbb{P})$ and the filtration $\mathbb{F}$ satisfying the usual conditions (Appendix Section 5.2).

Definition 1.3.1. *The H-valued SDE is defined on $[0,T]$ with the form*

$$dX(t)=(AX(t)+F(t,X))dt+B(t,X)dW_t, \tag{1.68}$$

with $X(0)=\xi$ where $A:D(A)\subset H\to H$ is the C_0 semigroup generator. We denote by $\{S(t),t\geq 0\}$ the corresponding generated operators. F and B are defined on

$$F:\Omega\times[0,T]\times C([0,T],H)\to H$$
$$G:\Omega\times[0,T]\times C([0,T],H)\to\mathcal{L}_2(K_Q,H).$$

Moreover, ξ is a $\mathcal{F}_0$-measurable H-valued random variable.

Note that we write Ω as part of the domain of F, G since we simply move the argument from X to F using the equality $F(t, X)(\cdot) = F(t, X(\cdot))$.

For Equation (1.68) to be well defined, we will assume for the rest of the chapter the following. As a technical condition, we require the F, B to be adapted to the sigma algebra generated by the cylinders with base $[0, t]$ which is the content of the following definition.

Definition 1.3.2. *For $\{f_i\}_{i\in\mathbb{N}} \in H$ with H a Hilbert space, we denote the Borel cylinders as*

$$C_{f_1,\dots,f_n} \coloneqq \{x \in H^\star : (\langle f_1, x\rangle, \dots, \langle f_n, x\rangle) \in B(\mathbb{R}^n)\}. \tag{1.69}$$

In case $y \in C([0,T], K)$, we define the cylinders with base over $[0, t]$ as

$$\xi_t^n \coloneqq \{y \in C(y(s_1), \dots, y(s_n)) \in U\}, \tag{1.70}$$

where $s_i \leq t$ for ever $i = 1, \dots, n$ and $U \subset H^n$ open.

Remark 1.3.3. *The significance of the above definition lies in the fact that the σ-algebra generated by this class of cylinder sets is the coarsest σ-algebra with respect to which all continuous linear functionals (on the dual space), or equivalently, all evaluations along finite collections of times or directions, are measurable. In particular, this σ-algebra ensures the measurability of all continuous functionals on path space, making it the natural σ-algebra for analyzing stochastic processes with continuous trajectories.*

Assumption 1.3.4.

- *F, B are jointly measurable and adapted with respect to $\mathcal{F}_t, \mathbb{C}_t$, where $\mathbb{C}_t$ is the σ algebra generated by cylinders with base $[0, t]$.*
- *F, B are jointly continuous*
- *F and B grow at most linearly in the whole trajectory, i.e., there exists a constant α such that for all $x \in C([0,T], H)$*

$$\|F(\omega, t, x)\|_H + \|B(\omega, t, x)\|_{\mathcal{L}_2(K_Q, H)} \leq \ell \left(1 + \sup_{0\leq s\leq T} \|x(s)\|_H\right) \tag{1.71}$$

  for $\omega \in \Omega$.

A crucial assumption for the existence and uniqueness of solutions is jointly Lipschitz, which is the subject of the next definition

Definition 1.3.5. *F, B are jointly Lipschitz if there exists an $M < \infty$ such that*

$$\|F(\omega, t, x) - F(\omega, t, y)\|_H + \|B(\omega, t, x) - B(\omega, t, y)\|_{\mathcal{L}_2(K_Q, H)} \leq M \sup_{0\leq s\leq T} \|x(s) - y(s)\|_H . \tag{1.72}$$

1.3.1 Solutions of SPDEs

Here we present two types of solutions, i.e., strong and mild. As the name suggests, the strong solution is the strongest of the two, which also implies that the conditions for such a solution to exist are much more restrictive.

Definition 1.3.6. *An adapted stochastic process on the filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\leq T}, \mathbb{P})$ *is a strong solution of the SDE in Equation* (1.68) *if the following conditions hold true*

- $X(\cdot) \in C([0,T], H)$
- $X(t,\omega) \in D(A)$, *$dt \otimes d\mathbb{P}$ almost everywhere*
- *the norms of the operators AX and $F + B$ are bounded almost surely, i.e.*

$$\mathbb{P}\left(\int_0^T \|AX(t)\|_H \, dt < \infty\right) = 1, \tag{1.73}$$

  and

$$\mathbb{P}\left(\int_0^T \left(\|F(t,X)\|_H + \|B(t,X)\|^2_{\mathcal{L}_2(K_Q,H)}\right) dt < \infty\right) = 1; \tag{1.74}$$

- *X is a solution of Equation* (1.68)*, i.e. for all $t \leq T$*

$$X(t) = \xi_0 + \int_0^t (AX(s) + F(s,X)) \, ds + \int_0^t B(s,X) \, dW_s \quad \mathbb{P}\text{-}a.s. \tag{1.75}$$

One of the biggest requirements that X has to satisfy is that $X(t,\omega) \in D(A)$ almost everywhere. We now present the concept of a mild solution, which relaxes this assumption.

Definition 1.3.7. *An adapted stochastic process on the filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\leq T}, \mathbb{P})$ *is a mild solution of the SDE in Equation* (1.68) *if*

$$\mathbb{P}\left(\int_0^T \|X(t)\|_H \, dt < \infty\right) = 1,$$
$$\mathbb{P}\left(\int_0^T \left(\|F(t,X)\|_H + \|B(t,X)\|^2_{\mathcal{L}_2(K_Q,H)}\right) dt < \infty\right) = 1;$$

and

$$X(t) = S(t)\xi_0 + \int_0^t S(t-s)F(s,X) \, ds + \int_0^t S(t-s)B(s,X) \, dW_s. \tag{1.76}$$

We can note that in Equation (1.76), the operator A is no longer applied to the process X, thus we remove the constraint that $X \in D(A)$, and since $S(t) \in L(H)$ is densely defined, we completely remove that constraint. The concept of mild solutions borrows from the variation of constants formula for the deterministic differential equation counterpart.

1.4 Application to forward pricing

In the context of forward pricing (Appendix, Chapter 5.6), the pricing functional is represented as an element in a function space. We will now outline the settings that will be used later.

For this section we will assume we work under the risk neutral probability, this motivates the change of notation from the probability measure $\mathbb{P}$ to $\mathbb{Q}$. Let $(\Omega, \mathcal{F}, \mathbb{Q})$ with the filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t \leq T}$ be a filtered probability space satisfying the usual conditions. We now assume $(H, \langle \cdot, \cdot \rangle_H)$ to be a separable Hilbert space as before, but now it is the space of measurable functions $f : \mathbb{R}^+ \to \mathbb{R}$ with associated induced norm $\| \cdot \|_H$. We denote the evaluation functionals $\delta_\xi \in H^\star$ for any $\xi \in \mathbb{R}^+$ defined as $\delta_\xi(f) = f(\xi)$. This implies that every $f \in H$, $\delta_\xi f = f(\xi)$ is continuous at ξ. However, since this is true for every $\xi \in \mathbb{R}^+$, we are restricting the functions $f \in H$ to be continuous. Furthermore, we assume $(S(t))_{t \geq 0}$ to be the family of right-shift operators defined by $S(t)f \coloneqq f(t + \cdot)$. This defines a strongly continuous (i.e. C_0) pseudo-contractive semigroup. We will make use of the following fact.

Lemma 1.4.1 (Chapter 1 of [28]). *The generator of the family $(S(t))_{t \geq 0}$ of right-shift operators, defined by*

$$(S(t)f)(\xi) \coloneqq f(t + \xi),$$

is the derivative operator ∂_ξ.

Proof. Straightforwardly, $\frac{(S_t f)(\xi) - f(\xi)}{t} = \frac{f(t+\xi) - f(\xi)}{t}$, which converges to $\partial_\xi f$ as $t \to 0$, by the definition of strong derivative. The density comes directly from the well-known fact that $C^\infty(\mathbb{R}^+)$ is dense in $L^p(\mathbb{R}^+)$ with $p \geq 1$, see [21] for the proof. $\square$

The final assumption is that H is a Banach algebra under pointwise multiplication of functions, which is satisfied in the Filipović space H_w for a choice of w.

When pricing forward options, the first choice to make is how to model the underlying asset price. As we shall see in practice later, this is generally modelled through the following SPDE.

$$dX(t) = \partial_\xi X(t)\, dt + \alpha(t, X(t))\, dt + \eta(t, X(t))\, dW(t) + \int_H \gamma(t, X(t), z)\, \widetilde{N}(dt, dz), \quad (1.77)$$

with $X(0) = X_0$ being an $\mathcal{F}_0$ measurable H-valued squared integrable random variable. Moreover, $(W_t)_{t \geq 0}$ is a Q-Wiener process with a positive trace-class operator Q, $\widetilde{N}(dt, dz) \coloneqq N(dt, dz) - dt \otimes v(dz)$ where $N(dt, dz)$ is a homogeneous Poisson random measure and the compensator $dt \otimes v(dz)$ with v being a σ-finite Lévy measure compensator. For the definition of Poisson measures and Lévy processes, consult Appendix 5.2.1. The compensator part makes the $\int_H \gamma(t, X(t), z) \widetilde{N}(dt, dz)$ a martingale. In other words, the number of jumps in a measurable set $\Delta t \times A \subset \mathbb{R}^+ \times H$ is given by the product $\Delta t \cdot \nu(A)$. Finally, the coefficients belong to the following classes: $\alpha : \mathbb{R}^+ \times H \to H$, $\eta : \mathbb{R}^+ \times H \to \mathcal{L}_2(H, H)$ and finally $\gamma : \mathbb{R}^+ \times H \to H$.

The final result of this section, after having introduced the notion of a mild solution of an SDE, is the existence and uniqueness of a mild solution to the SPDE with jumps given in Equation (1.77). Corollary 10.6 in [26] shows that, under standard Lipschitz and linear growth conditions on the coefficients, Equation (1.77) admits a unique mild solution. For completeness, we restate this result below.

Theorem 1.4.2 (Corollary 10.6 from [27]). *In the context of Equation* (1.77) *together with the filtered probability space* $(\Omega, \mathcal{F}, \mathbb{Q})$ *and the filtration* $\mathbb{F}$*, assume the following conditions:*

$$|\alpha(\cdot, 0)| \in L^p_{\text{loc}}(\mathbb{R}^+)41$$

$$\|\eta(\cdot, 0)\|_{\mathcal{L}_2(H)} \in L^p_{\text{loc}}(\mathbb{R}^+),$$

$$\int_H |\gamma(\cdot, 0, z)|^2 \, \nu(dz) \in L^1_{\text{loc}}(\mathbb{R}^+),$$

Moreover, suppose the following functions are Lipschitz continuous with respect to the spatial variable; that is, there exists a constant $C_t > 0$*, depending on* t*, such that*

$$|\alpha(t, x) - \alpha(t, y)| \leq C_t |x - y|,$$

$$\|\eta(t, x) - \eta(t, y)\|_{\mathcal{L}_2(H)} \leq C_t |x - y|,$$

and

$$\int_H |\gamma(t, x, z) - \gamma(t, y, z)|^2 \nu(dz) \leq C_t |x - y|.$$

Then there exists a unique, adapted to the filtration $\mathbb{F}$ *, mean-square continuous,* H*-valued càdlàg process* $X(t)$*, for* $t \geq 0$*, which satisfies the mild formulation of Equation* (3.13)*, i.e.,*

$$\begin{aligned} X(t) = \mathcal{S}_t X_0 + \int_0^t \mathcal{S}_{t-s} \alpha(s, X(s)) \, ds + \int_0^t \mathcal{S}_{t-s} \eta(s, X(s)) \, dW(s) \\ + \int_0^t \int_H \mathcal{S}_{t-s} \gamma(s, X(s-), z) \, \widetilde{N}(ds, dz), \end{aligned} \tag{1.78}$$

with $X(s-) \coloneqq \lim_{u \uparrow s} X(u)$*. Furthermore, we have*

$$\mathbb{E}\left[\sup_{0 \leq t \leq \tau} |X(t)|^2 \right] < \infty. \tag{1.79}$$

Having established the theoretical framework and assumptions underlying the stochastic partial differential equation (1.77), we will explore in subsequent chapters theoretical results and numerical methods appropriate for its practical resolution. In particular, we shall illustrate the infinite-dimensional Neural Network that will be used to solve this SPDE numerically. We first introduce the concept of Neural Network in the infinite-dimensional settings, further, we present a Universal Approximation Theorem in Theorem 2.9.

2 Neural Networks

In this section, we outline the fundamental mathematical definitions of Neural Networks, and their working mechanisms, first in the traditional real-valued cases, and then we extend them to the Hilbert-space-valued ones which are formalized in [9]. These tools will be crucial for understanding their application in the context of SDE, particularly in the specific case of forward pricing. Throughout this work, we will work exclusively with feedforward neural networks, which are introduced below.

2.1 Finite-dimensional Feedforward Neural Network

We start by introducing the definition of a real-valued neural network

Definition 2.1.1. *Let $\sigma : \mathbb{R} \to \mathbb{R}, a_i \in \mathbb{R}^n, l_i, b_i \in \mathbb{R}$, we define a single neuron to be the continuous map $\mathcal{N}_{l_i,a_i,b_i} \in C(\mathbb{R}^n, \mathbb{R})$ for $i = 1, \dots, N$ for $N < \infty$ defined as*

$$\mathcal{N}_{l_i,a_i,b_i}(x) := l_i \sigma(a_i^T x + b_i). \tag{2.1}$$

Furthermore, we define the corresponding neural network $\mathcal{N}_{l,a,b}$ to be

$$\mathcal{N}_{l,a,b}(x) := \sum_{i=1}^{N} \mathcal{N}_{l_i,a_i,b_i}(x), \tag{2.2}$$

for $a = (a_1, \dots, a_N), b = (b_1, \dots, b_N), l = (l_1, \dots, l_N)$.

The function σ takes the name *activation function*. The important result regarding feedforward neural networks is the Universal Approximation Theorem, where under certain conditions on σ, we have uniform convergence on a compact subset of $\mathbb{R}^n$.

Theorem 2.1.2 (Corollary 2.5 of [35])**.** *Assuming $\sigma : \mathbb{R} \to \mathbb{R}$ such that $\lim_{x\to\infty} \sigma(x) = 1$ and $\lim_{x\to-\infty} \sigma(x) = 0$. Then, for every $f \in C_b(\mathbb{R}^n, \mathbb{R})$, and for any $K \subset \mathbb{R}^n$ compact, there exist a sequence $l_i^{(n)}, a_i^{(n)}, b_i^{(n)}$ for $i = 1, \dots, N$, such that*

$$\sup_{x\in K} \left| f(x) - \mathcal{N}_{l^{(n)},a^{(n)},b^{(n)}}(x) \right| \to 0. \tag{2.3}$$

This is an extremely useful result, in fact, provided that we have a constructive procedure to find $l^{(n)}, a^{(n)}, b^{(n)}$, we can approximate any function on a compact. Recent work such as [48] proved the above result but in a much more general noncompact setting for neural networks with $n \geq 2$ hidden layers. However, this result will not be necessary for the rest of this work, as in the infinite dimensional settings, the compactness is still required.

In practice, there are several challenges, for instance, we don't have access to the function f but only to the so called "empirical measure". For a dataset $\mathcal{X} \in \mathbb{R}^{M \times n}, \mathcal{Y} \in \mathbb{R}^M$, for $x \in \mathbb{R}^n$, δ_x represents the Dirac measure at x, we have the estimate of f given by $\frac{1}{M}\sum_{i=1}^M \mathcal{Y}_i \delta_{\mathcal{X}_i}$ where $M > 1$ represents the dataset size and $\mathcal{X}_i$ is the i-th row of $\mathcal{X}$, thus $\mathcal{X}_i \in \mathbb{R}^n$. In the next section we analyze the technique that is commonly used to find the approximating sequence, although there are dozens of proposed methods and this area of research is in constant expansion.

Training

Given such expressive power of Neural Networks, we need to address the so-called *training* procedure, namely the procedure to compute the approximating sequence $l_i^{(n)}, a_i^{(n)}$, $b_i^{(n)}$. This is performed through the so-called gradient descent. Although for practical problems it generally leads to a correct minimizing sequence, currently there is no result guaranteeing that the corresponding sequence $\mathcal{N}_{l^{(n)},a^{(n)},b^{(n)}}$ converges in the sense of Equation (2.3).

We consider a dataset to be a list of pairs of the type $\{(x_j, f(x_j))\}_{j=1}^M$ for $x_j \in K$ sampled for a given distribution over K. Practically, we are interested in minimizing the so-called *loss* $L(l^{(n)}, a^{(n)}, b^{(n)}) := \sum_{i=1}^M \left|f(x_i) - \mathcal{N}_{l^{(n)},a^{(n)},b^{(n)}}(x_i)\right|^2$, by updating the current l_i^k, a_i^k, b_i^k, where k is our current step, by adding a scaled version of the gradient with respect to the loss. The loss is in fact the squared of the L^2 distance between the point $(f(\mathcal{X}_1), \dots, f(\mathcal{X}_M))$ and $(\mathcal{N}_{l^{(n)},a^{(n)},b^{(n)}}(x_1), \dots, \mathcal{N}_{l^{(n)},a^{(n)},b^{(n)}}(x_M))$. The algorithm below outlines the training procedure in full detail.

Algorithm 1 Gradient Descent for $\mathcal{N}_{l,a,b}$

Require: Dataset $\{(x_j, f(x_j))\}_{j=1}^M$, learning rate $\eta > 0$, iterations K

Initialize $l_i^{(0)}, a_i^{(0)}, b_i^{(0)}$ for $i = 1, \dots, N$
for $k = 0$ to $K-1$ **do**
    for $i = 1$ to N **do**
        Compute gradients:
          $g_{l_i} \leftarrow \nabla_{l_i} L\big(l^{(k)}, a^{(k)}, b^{(k)}\big)$
          $g_{a_i} \leftarrow \nabla_{a_i} L\big(l^{(k)}, a^{(k)}, b^{(k)}\big)$
          $g_{b_i} \leftarrow \nabla_{b_i} L\big(l^{(k)}, a^{(k)}, b^{(k)}\big)$
        Update parameters:
          $l_i^{(k+1)} \leftarrow l_i^{(k)} - \eta\, g_{l_i}$
          $a_i^{(k+1)} \leftarrow a_i^{(k)} - \eta\, g_{a_i}$
          $b_i^{(k+1)} \leftarrow b_i^{(k)} - \eta\, g_{b_i}$
    end for
end for

We refer to such an algorithm as *Stochastic Gradient Descent*. This is a vast area of research, and an extensive account of the subject is well beyond the scope of this work. We refer to [2], [58], and [30] for a more complete overview of the topic.

2.1.1 Beyond the compact case

Although not strictly necessary for the rest of the work, we include the following result for the sake of completeness on the topic. The universal approximation theorem has been recently extended to the noncompact case in [48] in the context of $\mathbb{R}^n$ with $n \in \mathbb{N}$. We will briefly discuss the main result here. Before stating the results, we need to introduce a slight variation of the notation in (2.2).

Definition 2.1.3. *Let $\sigma : \mathbb{R} \to \mathbb{R}$. For $n \in \mathbb{N}$ denoting the input nodes, we define the space of neural network with one hidden layer:*

$$\mathcal{N}^1(\mathbb{R}^n) \coloneqq \operatorname{span}\left\{x \to \sigma(a^T x + b) | a \in \mathbb{R}^n, b \in \mathbb{R}\right\}. \tag{2.4}$$

To extend the definition above to l hidden layers, we define the space of neural network recursively as follows:

$$\mathcal{N}^l(\mathbb{R}^n) \coloneqq \operatorname{span}\left\{x \to \sigma(f(x) + b) | f \in \mathcal{N}^{l-1}(\mathbb{R}^n), b \in \mathbb{R}\right\}. \tag{2.5}$$

We now state the universal approximation theorem for noncompact sets. The assumptions on the activation functions σ are relaxed.

Theorem 2.1.4 (Theorem 3.7 in [48])**.** *Let $n, l \in \mathbb{N}$ and the nonlinear activation function $\sigma \in C(\mathbb{R})$ be such that $\lim_{x\to\infty} \sigma(x) - a_1 x - b_1 = 0$ and $\lim_{x\to-\infty} \sigma(x) - a_2 x - b_2 = 0$ for some $a_1, b_1, a_2, b_2 \in \mathbb{R}$, then*

$$C_0(\mathbb{R}^n) \subseteq \overline{\mathcal{N}^l(\mathbb{R}^n)}. \tag{2.6}$$

The closure in the equation above is to be intended as usual in the topology induced by the supremum norm.

2.2 Hilbert-spaced valued

We now aim to generalize the results from the finite-dimensional real case to the infinite-dimensional setting, presenting results derived from [9]. This extension is motivated not solely by theoretical considerations but also by concrete practical applications. In particular, when the argument of the function f we seek to approximate is itself a function, the previously established framework and corresponding results become insufficient. For instance, solutions to partial differential equations typically cannot be represented adequately as points within a finite-dimensional space. Within the scope of this work, contexts such as the stochastic modeling of asset prices (e.g., the Black–Scholes model), stock price forecasting, option pricing, and hedging strategies naturally lead to infinite-dimensional domains. Since neural networks cannot be implemented in an actual infinite-dimensional space, we have to restrict the analysis to a compact subset. In particular, we will work with a finite-dimensional subspace spanned by finitely many basis vectors of H. Then, we will show the required conditions on the activation function σ so that the supremum convergence on compact sets still holds. It is worth mentioning the work of [14] [39] as data driven alternative approaches to the one presented here. Moreover, in [47], within the context of weighted spaces, the universal approximation theorem for noncompact subset was proven, we refer to the reference for more details.

Let the field $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and $(H, \langle\cdot,\cdot\rangle)$ be an $\mathbb{F}$-Hilbert space. We now extend the definitions of Neural Network from the real case to the space H. Clearly, the real case will be a special case of the Hilbert-space valued, thus, we will use the following definition throughout the rest of this work.

Definition 2.2.1. *Let $(H, \langle\cdot,\cdot\rangle)$ be a Hilbert space. Consider the continuous function $\sigma : H \to H$, called activation function. Consider the linear operator $A \in L(H)$, $b \in H$ and $l \in H^\star$. We define the neural network $\mathcal{N}_{l,A,b} \in C(H, \mathbb{F})$ as composition of continuous function*

$$x \to \langle l, \sigma(Ax + b)\rangle. \tag{2.7}$$

As in Equation (2.2), we are now interested in the linear combinations of the $\mathcal{N}_{l_i,a_i,b_i}$ for $i = 1, \dots, N$. Formally, we can represent it as $\sum_{i=1}^N \alpha_i \mathcal{N}_{l_i,a_i,b_i}$ for finite N. Note that we are slightly abusing the notation here since the $\mathcal{N}_{l_i,a_i,b_i}$ need not be the same for every linear combination. We define the corresponding vector subspace

$$\mathfrak{N}(\sigma) := \text{span}\,\{\mathcal{N}_{l,A,b}; l \in H^\star, A \in L(H), b \in H\} \tag{2.8}$$

The objective is then to prove the density of $\mathfrak{N}(\sigma)$ in $C(H, \mathbb{F})$ under a specific topology. The convergence in the supremum norm topology is yet to be prove; similarly to Theorem 2.1.2, we will restrict our study to the convergence in the supremum norm over compact sets only.

We endow $C(H, \mathbb{F})$ with the intuitive choice of supremum norm on compacts. Formally, we choose the family of seminorm (recall Definition 5.1.1 in the Appendix) $\mathcal{P}$ to be

$$\mathcal{P} := \{q_K(f); K \subset H \text{ compact}, f \in C(H, \mathbb{F})\}, \tag{2.9}$$

where $q_K(f) := \sup_{x\in K} |f(x)|$. We then consider the topology generated by this family of seminorms; this, by definition, makes $C(H, \mathbb{F})$ into a locally convex space. In this setup, we know that the dual of $C(H, \mathbb{F})$ is the space of measures on compacts. Namely, for $L \in C(H, \mathbb{F})^\star$ and $A \in C(H, \mathbb{F})$, then there exists a compact K and regular Borel measure μ on K such that $\int_K A d\mu = L(A)$ (Proposition 4.1 of [22]). To prove the convergence result, we need the following proposition and definition.

Proposition 2.2.2 (Proposition 4.1 in [22])**.** *If $\phi : C(H, \mathbb{F}) \to \mathbb{F}$ continuous and linear, then there exists a compact set $K \subset H$ and a regular Borel μ on K such that*

$$\phi(f) = \int_K f d\mu \tag{2.10}$$

for every $f \in C(H, \mathbb{F})$. .

Definition 2.2.3. *We call the continuous function $\sigma : H \to H$ discriminatory if for any fixed pair μ, K, where μ is a regular Borel measure on the compact set $K \subset H$,*

$$\int_K \langle l, \sigma(Ax+b)\rangle \mu(dx) = 0 \tag{2.11}$$

for all $l \in H^\star, A \in L(H), b \in H$ implies that $\mu = 0$.

This is reminiscent of the fundamental calculus of variation, which states that if for a continuous function f on an open interval (a,b), $\int_a^b f(x)h(x)dx = 0$ for every compact supported smooth function h, then $f = 0$. The concept of discriminatory function allows us to state and prove the following result, which is an essential theorem of this section.

Theorem 2.2.4 (Theorem 2.3 of [9])**.** *For H an $\mathbb{F}$-Hilbert space, let $\sigma : H \to H$ be continuous and discriminatory. Then $\mathfrak{R}(\sigma)$ is dense in $C(H, \mathbb{F})$ with the topology induced by the family of seminorms of the form as in Equation* (2.9)*.*

Proof. We start by noting that $\mathfrak{R}(\sigma) \subset C(H, \mathbb{F})$ since a linear combination of a continuous functions is continuous, in particular, $\mathrm{cl}(\mathfrak{R}(\sigma)) \subset C(H, \mathbb{F})$ since $C(H, \mathbb{F})$ is closed under the supremum topology. We argue by contradiction assuming that $\mathrm{cl}(\mathfrak{R}(\sigma)) \subsetneq C(H, \mathbb{F})$. The tactic is now to use the Hahn-Banach Theorem, together with the discriminatory property of σ to get a contradiction.

Consider u_0 belonging to $C(H, \mathbb{F}) \setminus \mathrm{cl}(\mathfrak{R}(\sigma))$. Note that $C(H, \mathbb{F}) \setminus \mathrm{cl}(\mathfrak{R}(\sigma))$ is open as a complement of a closed set, therefore, there exists an open U in $C(H, \mathbb{F}) \setminus \mathrm{cl}(\mathfrak{R}(\sigma))$ that contains u_0. Since the space is locally convex, we may choose U to be convex while remaining open and containing u_0. By one of the versions of Hahn-Banach Theorem (Theorem 3.7 of [22]), there exists a continuous linear functional $\phi : C(H, \mathbb{F}) \to \mathbb{F}$ with the property that $\phi\restriction_{\mathrm{cl}(\mathfrak{R}(\sigma))} = 0$ and $\mathfrak{R}(\phi) > 0$ on U and in particular ϕ is not trivially zero.

By Proposition 3.4 of [22], there exists a compact subset $K \subset H$ and a nonzero regular Borel measure such that for every $f \in C(H, \mathbb{F})$

$$\phi(f) = \int_K f(x)\mu(dx). \tag{2.12}$$

Now note that by the definition of the neural network, $\mathcal{N}_{l,A,b}$ lies in $C(H, \mathbb{F})$, thus, we can choose f to be a neural network. Thus

$$\int_K \langle l, \sigma(Ax+b)\rangle \mu(dx) = 0, \tag{2.13}$$

where $l \in H', A \in L(H), b \in H$. However, since σ is discriminatory, we have that $\mu = 0$, which is obviously a contradiction. $\square$

We were then able to obtain a similar result to Theorem 2.1.2 but for the Hilbert-valued neural network. However, the condition on σ is more restrictive than the one in Theorem 2.1.2; there are, however, several function classes that satisfy such conditions, making the result useful in practice and not only a purely theoretical result. For extensive examples of such a function, consult [9].

Besides compactness-based approaches, other directions have been explored in the literature. The line in [14] uses latent space projections via PCA [39], where infinite-dimensional input/output data are mapped to finite-dimensional subspaces, a neural network is trained there, and then lifted back; this yields convergence in $L^2(\mu)$ under suitable conditions. The alternative line in [23] employs weighted function spaces, where admissible weight functions render noncompact subsets effectively compact. Within this framework, universal approximation theorems have been proved for Banach-valued neural networks, allowing dense approximation in weighted function spaces.

3 Forward pricing

3.1 Filipović space

In this section, we devote our attention to the Filipović space, focusing on the results that are relevant to our case and trying to motivate the choice of this space as the function space used to model the solutions of the forward price functional. The content will contain results from [13] and [25].

We begin by recalling the following definition.

Definition 3.1.1. *Let I be an interval of $\mathbb{R}$. We call a function $f : I \to \mathbb{R}$ absolutely continuous on I if for every positive number ϵ, there exists $\delta > 0$ such that for every finite sequence of pairwise disjoint sub-intervals $(x_k, y_k) \subset I$ satisfy*

$$\sum_{i=1}^{N}(y_k - x_k) < \delta \tag{3.1}$$

then

$$\sum_{i=1}^{N}|f(y_k) - f(x_k)| < \epsilon. \tag{3.2}$$

We call the function f absolutely continuous if it is absolutely continuous on I for every $I \subset \mathbb{R}$.

We can see that Equation (3.2) is strictly related to the concept of total variation of f. If we consider the measure corresponding to f, that is $\mu_f([a, b]) = f(b) - f(a)$, the notion of f being absolutely continuous is equivalent to μ_f being absolutely continuous with respect to the Lebesgue measure. As it is well known, a measure that has such a property allows for the Radon-Nikodym derivative and the corresponding fundamental theorem of calculus.

The following proposition formalizes this.

Proposition 3.1.2 (Theorem 20.8 of [4])**.** *Let $f : \mathbb{R} \to \mathbb{R}$ and $[a, b]$ be a compact interval, then the following conditions are equivalent.*

- *f is absolutely continuous;*
- *f has a derivative f' defined a.e. and*

$$f(x) = f(a) + \int_a^b f'(t)dt, \tag{3.3}$$

for all x on $[a, b]$.

We are now ready to define the Filipović space.

Definition 3.1.3. *Let $w : \mathbb{R}^+ \to [1, \infty)$ be a non-decreasing measurable function with $w(0) = 1$. The Filipović space, denoted by H_w is the space of absolutely continuous functions $f : \mathbb{R}^+ \to \mathbb{R}$ such that*

$$|f|_w^2 := f^2(0) + \int_0^\infty w(x)|f'(x)|^2 dx < \infty. \tag{3.4}$$

H_w is clearly of linear vector space over $\mathbb{R}$. The following Lemma formalizes the applicability of the fundamental theorem of calculus to H_w.

Lemma 3.1.4 (Lemma 3.1 from [13])**.** *If $f \in H_w$, then $f' \in L^1([0, x])$ for any $x > 0$,*

$$|f'|_{L^1([0,x])} \leq \sqrt{x}\,|f|_w,$$

and

$$f(x) = f(y) + \int_y^x f'(z)\,dz,$$

for any $y \in \mathbb{R}_+,\ x \geq y$.

We define the inner product in H_w to be the following.

$$\langle f, g\rangle_{H_w} := f(0)g(0) + \int_0^\infty w(x)f'(x)g'(x)dx. \tag{3.5}$$

This lets us prove that the space H_w is indeed Hilbert as expected.

Proposition 3.1.5 (Proposition 5.1.1 of [25])**.** *The vector space H_w equipped with the norm $\|\cdot\|_{H_w}$ is a separable Hilbert space.*

Proof. We prove the assertion by showing that there exists an isometry from H_w to a separable Hilbert space; therefore, completeness and separability follow. Consider $U := \mathbb{R} \times L^2(\mathbb{R}^+)$ equipped with the norm $\left(|\cdot|^2 + \|\cdot\|^2_{L^2(\mathbb{R}^+)}\right)^{\frac{1}{2}}$. It is clear that U inherits the separable Hilbert space structure from $\mathbb{R}$ and $L^2(\mathbb{R}^+)$. We define then the following operator $T : H_w \to \mathbb{R} \times L^2(\mathbb{R}^+)$ by

$$Th = (h(0), h'w^{\frac{1}{2}}), \quad h \in H_w. \tag{3.6}$$

To show that $T(H_w) = U$ we provide the following inverse

$$\left(T^{-1}(u, f)\right)(x) = u + \int_0^x f(\eta)w^{-\frac{1}{2}}(\eta)\,d\eta, \quad (u, f) \in \mathbb{R} \times L^2(\mathbb{R}^+). \tag{3.7}$$

□

The Filipović space is a weighted L^2 space, and a standard choice for the weight function is $w(x) = \exp(\alpha x)$ with $\alpha > 0$ and $x \in \mathbb{R}^+$. With such w, the Filipović space supports the Nelson-Siegel curves [56], which are a popular choice for modelling forward rates in fixed-income theory.

An important property of the Filipović space is that every function is bounded given that w^{-1} is integrable.

Lemma 3.1.6 (Lemma 3.2 [13]). *Assume $w^{-1} \in L^1(\mathbb{R}^+)$, then*

$$\sup_{x\geq 0} |f(x)| \leq (1 + \sqrt{|w^{-1}|_{L^1(\mathbb{R}^+)}})|f|_w, \tag{3.8}$$

for any $f \in H_w$.

As we will see in the applications regarding forward pricing, the evaluation operator $\delta_x : H_w \to \mathbb{R}$ is the linear operator $\delta_x(f) = f(x)$ for $f \in H_w$.

It is easily verifiable that $\delta_x \in H_w$ by the following inequalities which follow from the fundamental theorem of calculus and Cauchy-Schwarz inequality applied to $\langle w^{\frac{1}{2}}, w^{\frac{1}{2}} f\rangle_{L^2}$.

$$\begin{aligned}
|\delta_x(f)| = |f(x)| &= |f(0) + \int_0^x f'(z)dz| \\
&\leq |f(0)| + \left| \int_0^x f'(z)dz \right| \\
&\leq |f(0)| + \left(\int_0^x w^{-1}(y)dy \right)^{\frac{1}{2}} \left(\int_0^x w(y)|f'(y)|^2 dy \right)^{\frac{1}{2}} \\
&\leq \left(1 + \left(\int_0^x w^{-1}(y)dy \right)^{\frac{1}{2}} \right) |f|_w < \infty.
\end{aligned}$$

In the context of forward option pricing, the functional δ_x is used as an element of H_w that, given the maturity t, returns the option price.

We dedicate the last portion of this section to the Banach algebra structure of the Filipović space. It turns out that, under a rescaling of the norm defined in 3.4, the Filipović space is indeed a Banach Algebra.

Proposition 3.1.7 (Proposition 3.1 from [56]). *Assume $w^{-1} \in L^1(\mathbb{R}^+)$, then, H_w is a Banach algebra with respect to the norm $\| \cdot \|_w := c| \cdot |_w$*
with $c = \sqrt{1 + 4\left(1 + \sqrt{|w^{-1}|_{L^1(\mathbb{R}^+)}}\right)^2}$.

As we will see in the application, the Banach Algebra structure allows the $\exp f$ to be in H_w for every $f \in H_w$. This is particularly useful as one of the possible modelling choices for the price functional could be $\exp(\sigma_x)$.

3.2 Price formulation

For the structure of the following sections, we will follow the outline of the paper [10]. We are interested in pricing forward price dynamics, i.e. the price of an asset in a forward contract. In short, a forward contract consists of two parties agreeing to exchange an underlying asset at a given price at a specified time in the future. Moreover, in the context of energy trading, fixed-deliver forwards are a specific type of forward contract where the two parties specify the exact amount of energy for a fixed period in time (see Appendix Section 5.6). Clearly, this type of forward contracts is easier to model than the flow forwards, where the energy delivery occurs over an extended period rather than on a specific date.

We now outline the problem more precisely; given a contract delivery period $[T_1, T_2]$ where $0 \leq T_1 < T_2$, the price at time $t \leq T_1$ of such forward contract is denoted by $\hat{F}(t, T_1, T_2)$. We can naturally extend this to the corresponding forward options, specifically estimating a call option price. From now on, we assume an arbitrage-free market and work under the corresponding martingale measure $\mathbb{Q}$. Given a strike price K and an exercise time $\tau \leq T_1$ the price of a call option on the forward contract is given by

$$V(t, \tau) = e^{-r(t-\tau)} \mathbb{E}\left[\max(\hat{F}(t, T_1, T_2) - K), 0 | \mathcal{F}_t\right]. \tag{3.9}$$

Where $r > 0$ is the risk-free interest rate. For simplicity and without loss of generality, in the following we will choose $r = 0$, moreover, in the actual numerics, we will focus on the most liquid options that are the ones with $\tau = T_1$.

This problem is generally formulated in terms of SPDEs (see Equation (3.13)), where the unknown process is $V(\cdot, \cdot)$. Due to their infinite-dimensional nature, numerical methods have their computational challenges. Some of these challenges are overcome in [10], where the aforementioned problem is recast to an optimization over continuous functionals, and using a density argument, it is shown that the minimizer can be obtained by optimizing just over the Lipschitz continuous functions, such as neural networks.

We now define the model for the underlying asset price, adopting the settings detailed in Section 1.4. For completeness, key aspects are reiterated here, but we refer to Section 1.4 for a comprehensive discussion.

3.3 SPDE formulation

First, let us recall the notion of the space of locally p integrable functions.

Definition 3.3.1. *Let a Lebesgue integrable function $f : \Omega \subset \mathbb{R} \to \mathbb{R}$. If f is such that for every compact $K \subseteq \Omega$ the following holds*

$$\int_K |f(x)|^p dx < \infty, \tag{3.10}$$

then f is called locally p integrable.

Remark 3.3.2. *It can be proven that an equivalent formulation of the definition of locally p integrable is that*

$$\int_K |f(x)\phi(x)|^p dx < \infty, \tag{3.11}$$

for each test function $\phi \in C_c^\infty(\Omega)$. In our context, when we state that $\alpha(\cdot, 0)$ is L^p_{loc} it simply means that for every finite $t \geq 0$

$$\int_0^t |\alpha(t,0)|^p dt < \infty. \tag{3.12}$$

The key idea about this definition is that the behavior at the boundary does not matter. In fact, it can be easily seen that the constant function 1 is L^p_{loc} but obviously not integrable in the L^p sense.

From this point forth, we adopt the standard setting we already encountered of a filtered probability space, and we work on a separable Hilbert space H of measurable functions on $\mathbb{R}_+$ with continuous evaluation functionals and a C_0 right–shift semigroup. Moreover, in the same fashion as in the previous section, we let $(\Omega, \mathcal{F}, (\mathcal{F})_t, \mathbb{Q})$ be a filtered probability space satisfying the usual conditions, which corresponds to the arbitrage-free measure. A canonical choice for the function space is the Filipović space H_w (absolutely continuous functions with $|x|_w^2 = x(0)^2 + \int_0^\infty w(\xi)\, x'(\xi)^2\, d\xi < \infty$), which is a Banach algebra under pointwise multiplication and admits continuous evaluations and a quasi-contractive shift semigroup. The underlying asset price is modeled by the following SPDE

$$dX(t) = \partial_\xi X(t)\, dt + \alpha(t, X(t))\, dt + \eta(t, X(t))\, dW(t) + \int_H \gamma(t, X(t), z)\, \widetilde{N}(dt, dz). \tag{3.13}$$

See Equation (1.77) in Section 1.4 and Appendix Section 5.2.1 for further details. Theorem 1.4.2 prove the existence of a mild solution to Equation (3.13), however, in our case, the settings are simplified by assuming that the coefficients α, η, γ are not state-dependent, i.e., they do not depend on $X(t)$, thus $\alpha(t, X(t)) = \alpha(t)$. In this case, the Lipschitz conditions in [26] are trivially satisfied. The existence of a mild solution of Equation (3.13) is thus given by

$$X(t) = \mathcal{S}_t Y_0 + \int_0^t \mathcal{S}_{t-s}\alpha(s)\, ds + \int_0^t \mathcal{S}_{t-s}\eta(s)\, dW(s) + \int_0^t \int_H \mathcal{S}_{t-s}\gamma(s,z)\, \widetilde{N}(ds, dz). \tag{3.14}$$

3.4 Forward price model

Let us move to introducing the price dynamics for a forward contract. Define $F(t,T)$ as the forward price at time $t \geq 0$ with deliver at time $T \geq t$. Suppose that $Y(t)$ satisfies Equation (3.14), then it holds true that $F(t,T) = Y(t, T-t)$. Then, the price dynamics can be defined by the continuous mapping $t \mapsto \exp(Y(t))$. Given the assumption that H is a Banach algebra, it follows that $X(t) \coloneqq \exp(Y(t)) \in H$. In fact, the exponential of an operator is defined as $\exp Y(t) = \sum_{n=0}^\infty \frac{Y(y)^n}{n!}$ for $t \geq 0$. Given that H is a Banach algebra, we directly see that $\left\| \frac{1}{n!} Y(t)^n \right\| \leq \frac{1}{n!} \|Y(t)\|^n$, therefore, the sum converges in norm for every $t \geq 0$, thus $\exp(Y(y)) \in H$.

Recall that, since we are working in a Hilbert function space, $Y(t)$ is a function $Y(t) : \mathbb{R}^+ \to \mathbb{R}$; thus, $Y(t, \xi) \coloneqq Y(t)(\xi)$ denotes the evaluation of the functional $Y(t)$ at ξ. Therefore, we have the following equivalences:

$$F(t, T) = X(t, T-t) = \delta_{T-t} X(t) = \delta_{T-t} \exp Y(t) = \exp Y(t, T-t). \tag{3.15}$$

For the flow forward, we look at the interval $[T_1, T_2]$ and consider the flow price for the interval as the average of the fixed price along this interval. Formally,

$$\hat{F}(t, T_1, T_2) \coloneqq \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} F(t, T) dT. \tag{3.16}$$

The following holds true

$$\begin{aligned}
\hat{F}(t, T_1, T_2) = \hat{F}(t, t+\xi, t+\xi+\lambda) &= \frac{1}{\lambda} \int_{t+\xi}^{t+\xi+\lambda} F(t, T) dt \\
&= \frac{1}{\lambda} \int_0^\infty \mathbb{1}_{[t+\xi, t+\xi+\lambda]}(T) F(t, T) dT \\
&= \frac{1}{\lambda} \int_0^\infty \mathbb{1}_{[t+\xi, t+\xi+\lambda]}(T) X(t, T-t) dT \\
&= \frac{1}{\lambda} \int_0^\infty \mathbb{1}_{[t+\xi, t+\xi+\lambda]}(u+t) X(t, u) du \\
&= \frac{1}{\lambda} \int_0^\infty \mathbb{1}_{[\xi, \xi+\lambda]}(u) X(t, u) du \\
&= \frac{1}{\lambda} \int_0^\infty \mathbb{1}_{[0, \lambda]}(u-\xi) X(t, u) du,
\end{aligned} \tag{3.17}$$

where we used the equivalence in Equation (3.15) and a change of variable $u \coloneqq T - t$. By setting $\xi \coloneqq T_1 - t$ and $\lambda \coloneqq T_2 - T_1$, we can rewrite Equation (3.17) as

$$\hat{F}(t, t+\xi, t+\xi+\lambda) = \frac{1}{\lambda} \int_0^\infty \mathbb{1}_{[0, \lambda]}(u-\xi) X(t, u) du, \tag{3.18}$$

by using the equality we just proved. We can therefore define the stochastic process

$$\hat{X}_\lambda(t)(t, \xi) \coloneqq \hat{F}(t, t+\xi, t+\xi+\lambda). \tag{3.19}$$

Moreover, we see $\hat{X}_\lambda(t)$ as the result of applying the integral operator $\mathcal{D}_\lambda$ to $X(t, u)$ as in Equation (3.18) with $\mathcal{D}_\lambda$ defined as follows

$$\mathcal{D}_\lambda(f) = \frac{1}{\lambda} \int_0^\infty \mathbb{1}_{[0, \lambda]}(u - \cdot) f(u) du. \tag{3.20}$$

3.5 The option price functional

In the previous Section, we looked at how to model the stochastic process of the underlying asset for the forward price. We now take a closer look at how to model option price functionals for forward contract delivering over the interval $[T_1, T_2]$. Here we assume that the option has the exercise time $\tau \leq T_1$ (the time at which the holder of the option can choose to exercise the right specified by the contract), and the payoff function is a measurable function $\mathfrak{P} : \mathbb{R} \to \mathbb{R}$. Using the results from [15], the price of the option is given by

$$V(t) := e^{-r(\tau-t)}\mathbb{E}\left[\mathfrak{P}\left(\widehat{F}\left(\tau, T_1, T_2\right)\right) \mid \mathcal{F}_t\right]. \tag{3.21}$$

Equivalently, using the notation introduced above, it can be rewritten as

$$V(t) = e^{-r(\tau-t)}\mathbb{E}\left[\mathfrak{P}\left(\delta_{T_1-\tau}\mathcal{D}_{T_2-T_1}(X(\tau))\right) \mid \mathcal{F}_t\right]. \tag{3.22}$$

Similarly to the specific case in Equation (3.9), for standard European call option, one can set $\mathfrak{B}(x) = \max(x - K, 0)$.

We now state an important proposition.

Proposition 3.5.1 (Proposition 4.1 from [10]). *Assume that X is given by Equation* (3.13) *and the coefficients α, η, γ are time independent. It then holds*

$$V(t,x) := e^{-r(\tau-t)}\mathbb{E}\left[\mathfrak{P}\left(\delta_{T_1-\tau}\mathcal{D}_{T_2-T_1}\left(X^{t,x}(\tau)\right)\right)\right] \tag{3.23}$$

with $X^{t,x}$ defined such that $X^{t,x}(t) = x \in H$.

Proof. Since the coefficients are independent of time by assumption, we can apply Theorem 9.30 from [51] to observe that X is Markovian. Thus,

$$\begin{aligned} e^{-r(\tau-t)}\mathbb{E}\left[\mathfrak{P}\left(\delta_{T_1-\tau}\mathcal{D}_{T_2-T_1}(X(\tau))\right) \mid \mathcal{F}_t\right] &= e^{-r(\tau-t)}\mathbb{E}\left[\mathfrak{P}\left(\delta_{T_1-\tau}\mathcal{D}_{T_2-T_1}(X(\tau))\right) \mid X(t)\right] \\ &= e^{-r(\tau-t)}\mathbb{E}\left[\mathfrak{P}\left(\delta_{T_1-\tau}\mathcal{D}_{T_2-T_1}(X^{t,X(t)}(\tau))\right)\right] \\ &= V(t,x). \end{aligned}$$

$\square$

This proposition allows us to see the pricing of the option as a functional of X. In order to apply the universal approximation theorem for neural network ([36], Proposition 3.3), we need the Lipschitz continuity of V which is exactly the content of the next proposition.

Proposition 3.5.2 (Proposition 4.3 from [10]). *Assuming $\mathfrak{P}$ is Lipschitz continuous, it follows that V is well defined and Lipschitz continuous.*

Proof. We begin by denoting with K the Lipschitz constant of $\mathfrak{B}$. Thus

$$|\mathfrak{P}(u)| - |\mathfrak{P}(0)| \leq |\mathfrak{P}(u) - \mathfrak{P}(0)| \leq K|u|. \tag{3.24}$$

It follows that

$$|\mathfrak{P}(\delta_{T_1-\tau}\mathcal{D}_{T_2-T_1}(X(\tau)))| \leq K\|\delta_{T_1-\tau}\|\|\mathcal{D}_{T_2-T_1}\||X(\tau)| + |\mathfrak{P}(0)|, \tag{3.25}$$

where the norms of $\delta_{T_1-\tau}$ and $\mathcal{D}_{T_2-T_1}$ are the operator norm. From Equation (1.79) in Theorem (1.79), we see that the right-hand side is therefore bounded a.s.

From the mild solution (Equation (3.14)), we find $Y(\tau)$ given $Y(t)$ for $\tau > t$ as

$$Y(\tau) = \mathcal{S}_{\tau-t}Y(t) + \int_t^\tau \mathcal{S}_{\tau-s}\alpha(s)\,ds + \int_t^\tau \mathcal{S}_{\tau-s}\eta(s)\,dW(s) + \int_t^\tau \int_H \gamma(s,z)\,\widetilde{N}(ds,dz). \tag{3.26}$$

Which we can rewrite as

$$Y(\tau) = S_{\tau-t}Y(t) + Z_{t,\tau}, \tag{3.27}$$

where

$$Z_{t,\tau} := \int_t^\tau \mathcal{S}_{\tau-s}\alpha(s)\,ds + \int_t^\tau \mathcal{S}_{\tau-s}\eta(s)\,dW(s) + \int_t^\tau \int_H \gamma(s,z)\,\widetilde{N}(ds,dz). \tag{3.28}$$

Note that by using the shift semigroup

$$\exp(S_{\tau-t}Y(t))(x) = \exp((S_{\tau-t}Y(t))(x)) = \exp(Y(t)(x+\tau-t)) = \exp(Y(t))(x+\tau-t). \tag{3.29}$$

Furthermore,

$$\exp(Y(t))(x+\tau-t) = S_{\tau-t}\exp(Y(t))(x) = S_{\tau-t}X(t)(x). \tag{3.30}$$

Let's analyze $X(t) = \exp Y(t)$ once again, where $Y(t)$ is give by Equation (3.14). We can write

$$X(t) = \exp(Y(t)) = \exp(S_tY_0)\exp(Y^0(t)), \tag{3.31}$$

where $Y^0(t)$ is $Y(t)$ with the initial condition $Y(0) = 0$. Moreover, we also get that

$$X(\tau) = \exp(S_{\tau-t}Y(t))\exp(Z_{t,\tau}). \tag{3.32}$$

By plugging $X(\tau)$ in the definition of D_λ (Equation (3.20)), and using the fact that $X(\tau) = \exp(Y(\tau))$, together with Equation (3.30), we get

$$\begin{aligned} D_{T_2-T_1}(X(\tau)) &= \frac{1}{T_2-T_1}\int_0^\infty \mathbb{1}_{[0,T_2-T_1]}(u-\cdot)X(\tau,u)du \\ &= \frac{1}{T_2-T_1}\int_0^\infty \mathbb{1}_{[0,T_2-T_1]}(u-\cdot)\exp(Y(\tau+u))\exp(Y^0(\tau,u))du. \end{aligned}$$

By applying $\delta_{T_1-\tau}$ to $D_{T_2-T_2}(X(\tau))$ we obtain

$$\delta_{T_1-\tau}\mathcal{D}_{T_2-T_1}(X(\tau)) = \frac{1}{T_2-T_1}\int_0^\infty \mathbf{1}_{[0,T_2-T_1]}(u-(T_1-\tau))\exp(Y(\tau+u))\exp(Y^0(\tau,u))\,du. \tag{3.33}$$

Using Proposition 6.3 in [56], by setting $\mu = 0$ where μ is the process that has to be local martingale in the mentioned proposition, we have arbitrage-free condition and as a consequence of $\mu = 0$, we have that $\mathbb{E}\left[\exp(Y^0(\tau,u))\right] = 1$.

For the Lipschitz continuity of V, and Fubini-Tonelli,

$$\begin{aligned}|V(t,x)-V(t,y)| &\leq K\mathbb{E}\left[\left|\delta_{T_1-\tau}\mathcal{D}_{T_2-T_1}\left(\mathcal{S}_{\tau-t}(x-y)e^{Z_{t,\tau}}\right)\right|\right] \\ &\leq K\|\delta_{T_1-\tau}\|\|\mathcal{D}_{T_2-T_1}\|\|X(\tau)\|\|S_{\tau-t}\|\mathbb{E}\left[\exp Z_{t,\tau}\right]|x-y|.\end{aligned}$$

Since all the quantities are bounded uniformly, we conclude that the operator $V(t)$ is Lipschitz. □

Having established these results, we will apply them to the approximation of the price functional. In particular, we will study how well we can approximate them with functions in $L^2(\mu)$ for the ad-hoc measure μ.

Approximation of the price functional

To compute the option price, we are interested in solving the following formulation

$$\overline{V}(x) \coloneqq \mathbb{E}\left[\mathcal{X}(x)\right] \tag{3.34}$$

with $\overline{V} : H \to \mathbb{R}$ and $\mathcal{X}$ is a general random variable on H. In the context of the preceding discussion, we define $\mathcal{X}(x) \coloneqq e^{-r(\tau-t)}\mathfrak{P}\left(\delta_{T_1-\tau}\mathcal{D}_{T_2-T_1}\left(X^{t,x}(\tau)\right)\right)$. The objective of this section is to establish the approximation of $\overline{V}(x) = \mathbb{E}[\mathcal{X}(x)]$ via neural networks. After introducing the measure μ, Lemma 3.5.3 demonstrates that $\overline{V}(x)$ is the optimal approximation of the random variable $\mathcal{X}(x)$ when minimizing the expected $L^2(\mu)$ error - namely $\mathbb{E}\left[\int_H |\mathcal{X}(x)-g(x)|^2\mu(dx)\right]$ - over functions $g \in L^2(\mu)$. This result justifies the strategy of learning an approximation of $\overline{V}(x)$ by minimizing an empirical version of this L^2 objective. Secondly, Proposition 3.5.2 establishes the Lipschitz continuity of $\overline{V}(x)$ (assuming the payoff function $\mathfrak{P}$ is Lipschitz continuous), which is a key regularity property. This property, combined with Universal Approximation Theorems for Neural Networks (Theorem (2.9)), ensures that neural networks form a dense function class capable of accurately approximating the function $\overline{V}(x)$ within suitable function spaces, such as $C(K,\mathbb{R})$ for compact $K \subset H$, or $L^2(H,\mu;\mathbb{R})$. To do so, we introduce a measure $\mu : \mathcal{B}(H) \to [0,\infty]$ with the following property

$$\mathbb{E}\left[\int_H \mathcal{X}^2(x)\mu(dx)\right] < \infty. \tag{3.35}$$

This is a technical condition that allows us to prove Lemma 3.5.3, since we use the bound on Equation (3.35) to prove Lipschitz continuity. There exists an equivalent formulation of this assumption, that is $\int_H \max(1,|x|^2)\mu(dx) < \infty$, that gives a sufficient condition on μ for the property (3.35). For sake of brevity, we refer to [10] Lemma 5.1 for the details. We now come to an important lemma that states that $\overline{V}$ is the global minimizer over $L^2(\mu)$.

Lemma 3.5.3 (Lemma 5.2 from [10])**.** *Given the current settings, it holds true that*

$$\overline{V}(\cdot) = \arg\min_{g\in L^2(\mu)} \mathbb{E}\left[\int_H |\mathcal{X}(x)-g(x)|^2\mu(dx)\right]. \tag{3.36}$$

Proof. Take $g \in L^2(\mu)$, then

$$
\begin{aligned}
&\mathbb{E}\left[\int_H |\mathcal{X}(x) - g(x)|^2 \mu(dx)\right] = \\
&\quad = \mathbb{E}\left[\int_H \mathcal{X}^2(x) - 2\mathcal{X}(x)g(x) + g^2(x)\mu(dx)\right] \\
&\quad = \mathbb{E}\left[\int_H \mathcal{X}^2(x)\mu(dx)\right] - 2\mathbb{E}\left[\int_H \mathcal{X}(x)g(x)\mu(dx)\right] + \int_H g^2(x)\mu(dx) \\
&\quad = \mathbb{E}\left[\int_H \mathcal{X}^2(x)\mu(dx)\right] - 2\int_H \overline{V}(x)g(x)\mu(dx) + \int_H g^2(x)\mu(dx) \qquad \text{(by Fubini-Tonelli)} \\
&\quad = \mathbb{E}\left[\int_H \mathcal{X}^2(x) - V^2(x)\mu(dx)\right] - 2\int_H \overline{V}(x)g(x)\mu(dx) + \int_H g^2(x)\mu(dx) + \int_H \overline{V}^2(x)\mu(dx) \\
&\quad = \int_H \mathbb{E}\left[\left(\mathcal{X}^2 - \mathbb{E}\left[\mathcal{X}(x)\right]\right)^2\right]\mu(dx) + \int_H |\overline{V}(x) - g(x)|^2\mu(dx) \\
&\quad = \int_H \mathrm{Var}\left[\mathcal{X}(x)\right]\mu(dx) + \int_H |\overline{V}(x) - g(x)|^2\mu(dx)
\end{aligned}
$$

Thus

$$
\begin{aligned}
\arg\min_{g\in L^2(\mu)} \mathbb{E}&\left[\int_H |\mathcal{X}(x) - g(x)|^2\mu(dx)\right] = \\
&= \arg\min_{g\in L^2(\mu)} \int_H \mathrm{Var}\left[\mathcal{X}(x)\right]\mu(dx) + \int_H |\overline{V}(x) - g(x)|^2\mu(dx) \\
&= \arg\min_{g\in L^2(\mu)} \int_H |\overline{V}(x) - g(x)|^2\mu(dx) \\
&= \overline{V}(\cdot)
\end{aligned}
$$

□

This result is crucial to be able to have convergence of neural network approximations and to guarantee that it will converge to the global minima using $L^2(\mu)$ distance as minimizing objective.

Convergence of NN for infinite dimensional case

We now look at how we can use the results presented in Section 2.2 regarding Hilbert-space valued Neural network. In particular, we look at a specific type of activation function σ that satisfies the so-called separating property, which is introduced below. This allows us to show that for any activation function σ with such a property, it is also discriminatory (Definition 2.2.3). Together with Theorem (2.9), we establish uniform convergence on compacts as the subject of Proposition 3.5.7. For the required knowledge for this chapter, we refer to Section 2.

Definition 3.5.4 (Separating property)**.** *Let $\sigma : H \to H$ be a continuous function. We say that σ has the* separating property *if there exist a nonzero continuous linear functional $\phi \in H^* \setminus \{0\}$ and vectors $u_+, u_-, u_0 \in H$, such that either $u_+ \notin \operatorname{span}\{u_0, u_-\}$ or $u_- \notin \operatorname{span}\{u_0, u_+\}$, and the following limits hold:*

$$\lim_{\lambda \to \infty} \sigma(\lambda x) = \begin{cases} u_+, & \textit{if } x \in \Psi_+, \\ u_-, & \textit{if } x \in \Psi_-, \\ u_0, & \textit{if } x \in \Psi_0, \end{cases}$$

where

$$\Psi_+ = \{x \in H \mid \langle \phi, x \rangle > 0\}, \quad \Psi_- = \{x \in H \mid \langle \phi, x \rangle < 0\}, \quad \Psi_0 = \ker(\phi).$$

Note that a particular case of such property can be achieved by setting $u_0 = u_- = 0$ and $u_+ \neq 0$.

We briefly recall the definition of a Neural Network.

Definition 3.5.5. *For $l \in H^*$, $A \in L(H)$, $b \in H$, and an activation function $\sigma : H \to H$ we define the one layer neural network as*

$$\mathcal{N}_{l,A,b}(x) \coloneqq \langle l, \sigma(Ax + b) \rangle. \tag{3.37}$$

Furthermore, we define its range

$$\mathfrak{R}(\sigma) \coloneqq \operatorname{span}\{\mathcal{N}_{l,A,b} | l \in H^*, A \in L(H), b \in H\}. \tag{3.38}$$

A fundamental question is under what conditions and topology we have the density of $\mathfrak{R}(\sigma)$ in $C(H, \mathbb{R})$.

We now state a result connecting the notion of separating property (Definition 3.5.4) and discriminating property (Definition 2.2.3).

Theorem 3.5.6 (Theorem 2.8 of [9])**.** *Let H be a Hilbert space, $\sigma : H \to H$ be a continuous and bounded, and satisfying the separating property, then σ is discriminatory.*

Therefore, by Theorem (2.9), we have density of $\mathfrak{R}(\sigma)$ in $C(H, \mathbb{R})$. We have thus established that functions in $C(H, \mathbb{R})$ can be approximated arbitrarily well by neural networks employing a separating activation function σ, without requiring the network itself to be finite-dimensional. However, for any practical application, the neural network must be finite-dimensional. The following proposition addresses this constraint and constitutes a crucial step in approximating Hilbert space-valued processes using such finite-dimensional neural networks.

Proposition 3.5.7 (Proposition 4.1 from [9]). *For each $N \in \mathbb{N}$, let $\Pi_N : H \to \{e_1, \dots, e_N\}$ be the orthogonal projection to the first N orthonormal basis. Let $\sigma : H \to H$ be Lipschitz. Fix $f \in C(H, \mathbb{R})$, $K \subset H$ compact and $\epsilon > 0$. Furthermore, consider the following function*

$$\mathcal{N}^\epsilon := \sum_{j=1}^{M} \langle l_j, \sigma(A_j x + b_j)\rangle, \quad x \in H. \tag{3.39}$$

with $l_j \in H^*, A_j \in L(H)$, $b_j \in H$ and $M > 0, M \in \mathbb{N}$. Recall that $\langle l_j, \sigma(A_j x + b_j)\rangle$ is the classical dual evaluation of l_j and not the inner product. Moreover, assume that

$$\sup_{x \in K} |f(x) - \mathcal{N}^\epsilon(x)| < \epsilon. \tag{3.40}$$

It follows that for every $\delta > 0$, there exists $N_* = N(\mathcal{N}^\epsilon, \delta) \in \mathbb{N}$ such that for every $N \geq N_*$ the following holds

$$\sup_{x \in K} \left| f(x) - \sum_{j=1}^{M} \langle l_j \circ \Pi_N, \sigma\left(\Pi_N A_j \Pi_N x + \Pi_N b_j\right)\rangle \right| < \varepsilon + \delta. \tag{3.41}$$

Proof. Let $j \in \{1, \dots, M\}$, $N \in \mathbb{N}$, and $x \in K$. By the triangle inequality and the linearity of the inner product, we have

$$\begin{aligned}
&|\langle l_j,\, \sigma(A_j x + b_j)\rangle - \langle l_j \circ \Pi_N,\, \sigma(\Pi_N A_j \Pi_N x + \Pi_N b_j)\rangle| \\
&\leq |\langle l_j,\, \sigma(A_j x + b_j) - \Pi_N \sigma(A_j x + b_j)\rangle| \\
&\quad + |\langle l_j,\, \Pi_N \sigma(A_j x + b_j) - \Pi_N \sigma(\Pi_N A_j \Pi_N x + \Pi_N b_j)\rangle| \,.
\end{aligned}$$

Note that, since Π_N is an orthogonal projection, its operator norm is one for every N. Applying the triangle inequality again yields

$$\begin{aligned}
&|\langle l_j,\, \sigma(A_j x + b_j) - \Pi_N \sigma(A_j x + b_j)\rangle| \\
&\quad + |\langle l_j,\, \Pi_N \sigma(A_j x + b_j) - \Pi_N \sigma(\Pi_N A_j \Pi_N x + \Pi_N b_j)\rangle| \\
&\leq \|l_j\| \, \|\sigma(A_j x + b_j) - \Pi_N \sigma(A_j x + b_j)\| \\
&\quad + \|l_j\| \, \|\Pi_N \sigma(A_j x + b_j) - \Pi_N \sigma(\Pi_N A_j \Pi_N x + \Pi_N b_j)\| \\
&\leq \|l_j\| \, \|\sigma(A_j x + b_j) - \Pi_N \sigma(A_j x + b_j)\| \\
&\quad + \|l_j\| \, \|\sigma(A_j x + b_j) - \sigma(\Pi_N A_j \Pi_N x + \Pi_N b_j)\|.
\end{aligned}$$

We now analyze the second term. By the Lipschitz continuity of σ (with Lipschitz constant K_σ), it follows that

$$\begin{aligned}
&\|l_j\| \, \|\sigma(A_j x + b_j) - \sigma(\Pi_N A_j \Pi_N x + \Pi_N b_j)\| \\
&\leq \|l_j\| K_\sigma \|A_j x + b_j - \Pi_N A_j \Pi_N x - \Pi_N b_j\| \\
&\leq \|l_j\| K_\sigma \big(\|A_j x - \Pi_N A_j x\| + \|\Pi_N A_j x - \Pi_N A_j \Pi_N x\| + \|b_j - \Pi_N b_j\|\big) \\
&\leq \|l_j\| K_\sigma \big(\|A_j x - \Pi_N A_j x\| + \|A_j x - A_j \Pi_N x\| + \|b_j - \Pi_N b_j\|\big) \\
&\leq \|l_j\| K_\sigma \big(\|A_j x - \Pi_N A_j x\| + \|A_j\| \, \|x - \Pi_N x\| + \|b_j - \Pi_N b_j\|\big) \\
&\leq \|l_j\| K_\sigma \left(\sup_{x \in A_j K} \|x - \Pi_N x\| + \|A_j\| \sup_{x \in K} \|x - \Pi_N x\| + \|b_j - \Pi_N b_j\| \right).
\end{aligned}$$

Since K is compact, by the approximation property of the basis (i.e., $\Pi_N \to I$ strongly as $N \to \infty$), and by the boundedness of $\|l_j\|$ and K_σ, we can make

$$|\langle l_j, \, \sigma(A_j x + b_j)\rangle - \langle l_j \circ \Pi_N, \, \sigma(\Pi_N A_j \Pi_N x + \Pi_N b_j)\rangle|$$

arbitrarily small by choosing N sufficiently large. $\square$

To conclude this subsection, we want to show that we can minimize $\overline{V}(\cdot)$ not by using functions $g \in L^2(\mu)$ as stated in Theorem 3.5.3, but by neural networks. To achieve that, we need one final Lemma about the Lipschitz continuity of the map $g \to \mathbb{E}\left[\int_H |\mathcal{X}(x) - g(x)|^2 \mu(dx)\right]$.

Lemma 3.5.8 (Lemma 5.3 in [10]). *Let $\mu : \mathcal{B}(H) \to [0, \infty]$ be a measure, then the map $L^2(\mu) \to \mathbb{R}$ denoted by*

$$g \to \mathbb{E}\left[\int_H |\mathcal{X}(x) - g(x)|^2 \mu(dx)\right] \tag{3.42}$$

is locally Lipschitz continuous.

Proof. We start by noting the following elementary equality for $a, b, c \in \mathbb{R}$.

$$|a - b|^2 - |a - c|^2 = 2a(c - b) + (b + c)(b - c). \tag{3.43}$$

For $g, h \in L^2(\mu)$ we get

$$\begin{aligned}
|I(g) - I(h)| &= \left| \mathbb{E}\left[\int_H |\mathcal{X}(x) - g(x)|^2 \mu(dx) - \int_H |\mathcal{X}(x) - h(x)|^2 \mu(dx) \right] \right| \\
&= \left| \mathbb{E}\left[\int_H |\mathcal{X}(x) - g(x)|^2 - |\mathcal{X}(x) - h(x)|^2 \mu(dx) \right] \right| \\
&\leq 2\mathbb{E}\left[\int_H |\mathcal{X}(x)||g(x) - h(x)|\mu(dx) \right] + \int_H |g(x) + h(x)||g(x) - h(x)|\mu(dx) \\
&\hspace{20em} \text{By (3.43) and triangle inequality} \\
&\leq 2\mathbb{E}\left[\int_H \mathcal{X}(x)^2 \mu(dx) \right]^{\frac{1}{2}} \left(\int_H |g(x) - h(x)|^2 \mu(dx) \right)^{\frac{1}{2}} + \\
&+ \left(\int_H |g(x) + h(x)|^2 \mu(dx) \right)^{\frac{1}{2}} \left(\int_H |g(x) - h(x)|^2 \mu(dx) \right)^{\frac{1}{2}} \\
&\hspace{20em} \text{by Cauchy-Schwarz} \\
&= K \left(\int_H |g(x) - h(x)|^2 \mu(dx) \right)^{\frac{1}{2}} = K \|g(x) - h(x)\|_{L^2(\mu)}.
\end{aligned}$$

In the last step, we used the assumption in (3.35). □

Take now a compact $K \subset H$ and without loss of generality assume that $\operatorname{supp} \mu = K$ and $\mu(K) = 1$ so that it becomes a probability measure with compact support. Note that by the definition of neural network in Hilbert space, if σ is Lipschitz, so is the corresponding neural network. From Proposition 3.5.7, we have that $CL^2_{lip}(\mu)$ denoting the L^2 integrable w.r.t. μ and Lipschitz continuous functions, are contained in $\overline{\mathfrak{R}(\sigma)}$ with the closure taken in the topology of uniform convergence on compacts. Thus, the important conclusion from this section is that we can approximate $\overline{V}$ arbitrarily well using a finite-dimensional neural network.

The objective function for $\overline{V}(\cdot) = \arg\min_{g \in L^2(\mu)} \mathbb{E}\left[\int_H |\mathcal{X}(x) - g(x)|^2 \mu(dx) \right]$ can thus be approximated by

$$\inf_{g \in L^2(mu)} I(g) = \inf_{g \in \mathfrak{P}(\sigma)} \mathbb{E}\left[\int_H |\mathcal{X}(x) - g(x)|^2 \mu(dx) \right]. \tag{3.44}$$

3.5.1 Numerical method for infinite-dimensional forward pricing

Theoretical setup

The objective of this section is setting up the framework that we will be using for the experiments. We will introduce the space that is used in practice in [10], as well as other spaces that we use as comparison to gether further experimental evidence. We begin this section by taking a closer look at the evaluation operator. Although this operator has been introduced in the Notation section, the following remark is important and what discussed therein is used later in this chapter.

Remark 3.5.9. *The δ operator represents the point evaluation, i.e. $\delta_\xi(x) := \langle x, \xi\rangle := x(\xi)$. This is a bounded linear function, thus, by the Riesz representation Theorem, there exists a unique $h_\xi \in H_w$ such that $\delta_\xi = \langle h_\xi, x\rangle$ for all $x \in H_w$. Thus, $h_\xi(u) = \delta^*_\xi 1(u)$ for all $u \in \mathbb{R}^+$, and 1 being the constant function and δ^* being the adjoint operator. This can be seen in the following:*

$$\langle h_\xi, u\rangle = \langle \delta^\star_\xi 1, u\rangle = \langle 1, \delta_\xi u\rangle = \delta_\xi(u) = u(\xi). \tag{3.45}$$

where, with an abuse of notation, we use interchangeably the $\langle\cdot,\cdot\rangle$ symbol for both the inner product and evaluation functional but as pointed out, in the light of the Riesz representation theorem, we can treat linear functional application as the inner product of its representative.

Provided the measurability of the weight function w^{-1} for H_w , we obtain that the differential operator ∂_ξ generates the shift C_0 semigroup $(S_t)_{t\geq 0}$ which is semi-contractive [11].

The authors of [10] introduce the following basis for H_w. For $i = 1$, $\widetilde{e}_1(\tau) = 1$, and for $i > 1$

$$\tilde{e}_i(\xi) = \xi^{(i-2)}\exp(-\xi). \tag{3.46}$$

Proposition 3.5.10. *Let H_w be the Filipović space for w non-decreasing, measurable and $w(0) = 1$. For $i \in \mathbb{N}$, let $\widetilde{e}_i : \mathbb{R}^+ \to \mathbb{R}$ as in Equation (3.46), then $\{\widetilde{e}_i\}_{i\in\mathbb{N}}$ forms a basis of H_w.*

Proof. Take an arbitrary finite subset $I := \{i_n\}_{n=1}^N \subseteq \mathbb{N}$ with each $i_n \geq 1$. We first prove that $\hat{e}_i$ are linearly independent; we then show that if a linear combination is 0, the coefficients a_i must be zero. If $1 \in I$, then we can write the linear combinations as $a_1 + \sum_{n=2}^N a_{i_n}\xi^{i_n-2}e^{-\xi} = 0$. However, for large enough ξ , the term e^ξ dominates the polynomial ones, it follows that $a_1 = 0$. Without loss of generality we can therefore remove $\widetilde{e}_1$ from any linear combination. The remaining terms are $\sum_{n=2}^N a_{i_n}\xi^{i_n-2} = 0$. Trivially, for a polynomial to vanish on the whole $\mathbb{R}^+$ it must be that all the coefficients are 0.

We now prove that they are a dense subset of H_w. Consider the polynomials $L = \{\xi^k\}_{k\in\mathbb{N}}$. By [3], L is dense in $L^2(\lambda)$. Moreover, they are dense in $C(\mathbb{R}^+) \cap L^2(\lambda)$ using still the norm 3.4. Furthermore, by considering the weighted measure $e^{xi} d\lambda$, and since $L^2(e^\xi d\lambda) \subset L^2(\lambda)$, we have the density of $\{\widetilde{e}_i\}_{i\in\mathbb{N}}$ in $C(\mathbb{R}^+) \cap L^2(e^\xi d\lambda)$ as well using the subspace norm inherited from $L^2(\mu)$. Let $f \in H_w$ where in this case $w = e^\xi$ and set $g := f'$. Consider $g_n \in \operatorname{span}\{\xi^k e^{-\xi}\}$ such that $g_n \to g$ in $L^2(e^\xi d\lambda)$. By setting $f_n := f(0) + \int_0^\xi g_n(u)du$, the density follows from the fact that

$$\|f - f_n\|^2_{H_w} \leq \int_0^\infty \|g(\xi) - g_n(\xi)\|^2 e^\xi d\xi \to 0 \quad \text{as } n \to \infty. \tag{3.47}$$

□

Since the basis given in Equation (3.46) is not orthonormal, we apply the Gram-Schmidt orthonormalization procedure to construct an associated orthonormal basis. Furthermore, since only a finite number of basis functions will be used in the numerical implementation, we restrict our attention to the first ten elements of this orthonormalized system, which are presented in the Appending in Equation (5.33).

We consider the exponential model in Equation (3.15). We recall that the spot price of the forward delivery at time t with delivery at time T is defined as

$$F(t,T) = \delta_{T-t} Y(t) = \delta_{T-t} \exp X(t) \tag{3.48}$$

where the process $X(t)$ follows Equation (3.14). Moreover, the standard call option payoff function is considered, i.e.

$$\mathfrak{P}(x) := \max(x - K, 0), \tag{3.49}$$

the price is defined as in the previous section as $\overline{V}(\mathcal{X}) = \mathbb{E}\left[\mathcal{X}(x)\right]$ with $\mathcal{X}$ defined as

$$\mathcal{X}(x) = e^{-r(\tau-t)} \mathfrak{P}\left(\delta_{T_1-\tau} \mathcal{D}_{T_2-T_1}\left((\mathcal{S}_{\tau-t} x) \exp\left(Z_{t,\tau}\right)\right)\right) \tag{3.50}$$

following Equation (3.22).

Generalization of the basis construction

We now aim at describing a procedure to construct a basis of H_w starting from a basis of $L^2(\mathbb{R}^+)$. This allows us to use well known basis of $L^2(\mathbb{R}^+)$, which is indeed a separable Hilbert space, and map them into the context of the Filipović space. We begin with the construction of an isometric isomorphism.

Lemma 3.5.11. *Let $w : \mathbb{R}^+ \to [1, \infty)$ be a non-decreasing measurable function with $w(0) = 1$. The map $J : H_w \to \mathbb{R} \times L^2(\mathbb{R}^+)$ defined as*

$$J(x) := \left(x(0), \sqrt{w}(\cdot) x'(\cdot)\right) \qquad x \in H_w \tag{3.51}$$

is an isometric isomorphism.

Proof. The linearity of J is trivial. The inner product is preserved as for every $x, y \in H_w$

$$\begin{aligned}\langle J(x), J(y)\rangle_{\mathbb{R}\times L^2(\mathbb{R}^+)} &= \langle x(0)y(0)\rangle + \langle \sqrt{w}x', \sqrt{w}y\rangle_{L^2(\mathbb{R}^+)} \\ &= x(0)y(0) + \int_0^\infty w(\xi)x'(\xi)y'(\xi)d\xi = \langle x, y\rangle_{H_w}.\end{aligned}$$

Thus J is an isometry. Being linear, we only need to show that $\ker J$ is only the zero functional. This is in fact true since $x(0) = 0$ and $\sqrt{w}x' = 0$ almost surely. Since x is absolutely continuous, x must be the zero functional. Finally, to show the surjectivity of J we provide J^{-1}. For every $(a, h) \in \mathbb{R}\times L^2(\mathbb{R}^+)$, $J^{-1}((a, h))(\xi) := x(\xi) := a + \int_0^\xi \frac{h(u)}{\sqrt{w(u)}}du$ for every $\xi \geq 0$. Note that $x(0) = a$. We now show that x is absolutely continuous. For every $R > 0$,

$$\begin{aligned}\int_0^R \frac{h(u)}{\sqrt{w(u)}}du &\leq \int_0^R \left|\frac{h(u)}{\sqrt{w(u)}}\right| du \\ &= \left(\int_0^R h(u)^2 du\right)^{\frac{1}{2}} \left(\int_0^R \frac{1}{w(u)}du\right)^{\frac{1}{2}} \\ &\leq \|h\|_{L^2(\mathbb{R}^+)} \frac{R^1}{2} < \infty,\end{aligned}$$

where we used Hölder's inequality and the fact that $\frac{1}{w(u)} \leq 1$ since $w(u) \geq 1$ for any u. Moreover,

$$\int_0^\infty w(u)|x'(u)|du = \int_0^\infty |h(u)|^2 du = \|h\|^2_{L^2(\mathbb{R}^+)} < \infty.$$

□

Trivially, J^{-1} is an isometry as well. The consequence of this theorem is that we can now simply construct an ONB of H_w starting from an ONB of $L^2(\mathbb{R}^+)$.

As part of our experiments, we decide to run numerical tests using the Laguerre functionals, which are a weighted Laguerre polynomials, to guarantee that they are in L^2. Given the explicit series expansion of the Laguerre polynomials

$$L_n(\xi) := \sum_{k=0}^n \binom{n}{k} \frac{(-\xi)^k}{k!}, \tag{3.52}$$

we define the corresponding Laguerre functionals as follows

$$l_n(\xi) := e^{-\frac{\xi}{2}} L_n(\xi). \tag{3.53}$$

The Laguerre functionals are a complete orthonormal basis of $L^2(\mathbb{R}^+)$ [57].

Using Lemma 3.5.11, we can write explicitly the ONB for H_w. In particular, we choose $w(\xi) = e^\xi$. Then an ONB of H_w is

$$\{1\} \cup \left\{ \Phi_n(\xi) := \int_0^\xi \frac{l_n(u)}{\sqrt{w(u)}}du = \int_0^\xi e^{-u} L_n(u)du \quad \right\}_{n\in\mathbb{N}}. \tag{3.54}$$

We don't need to perform any orthonormalization as the functionals are already orthonormal and the isometric isomorphism J guarantees these property hold true for Φ_n as well. We report below the first 10 basis elements explicitly, which, in the same fashion as the one above, will form the basis for the compact subspace in the numerical experiments. In Section 5.7.1 we report the first 10 orthonormal basis.

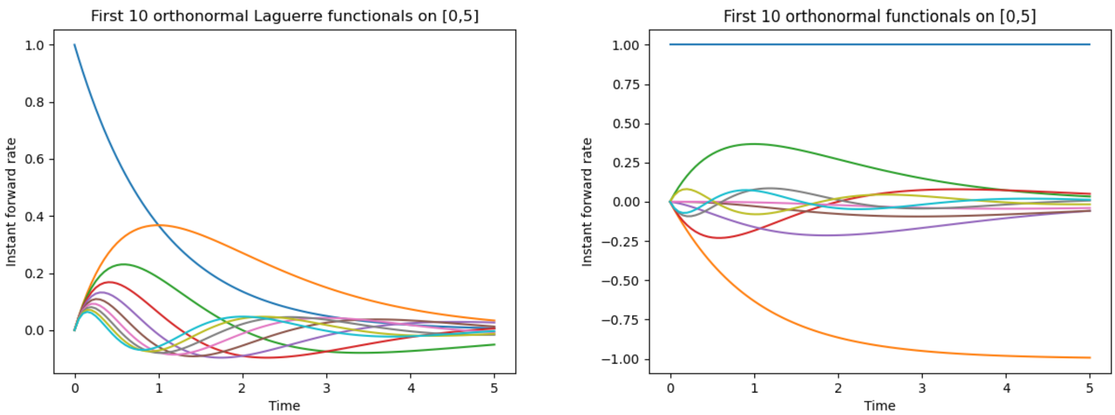

Figure 3.1: Plot of the Laguerre functionals obtained with the procedure described above and the basis proposed in [10].

4 Results

4.1 Numerical experiments

Setup
In these experiments we focus on monthly flow forward with the exercise time at the delivery period, namely $\tau = T_1$, moreover, we consider the period unit to be one year, namely $T_1 = \frac{1}{12}$ and $T_2 = \frac{2}{12}$. We aim to price the option at the current time $t = 0$ and without loss of generality, we set the strike price $K = 1$ and risk-free interest rate $r = 0$.

As mentioned, we are interested in approximating $\overline{V} : H_w \to \mathbb{R}$ to guarantee the convergence of the neural network, we need to restrict our attention to a compact subset of H_w. We consider the compact subset generated by the 7 orthonormal vectors mentioned in the previous section, with coefficients ranging in $[0, 1]$, formally

$$K_7 \coloneqq \{x \in H \mid x = \sum_{i=1}^{7} a_i e_i, \text{ where } a_i \in [0, 1]\}. \tag{4.1}$$

Finally, the last ingredient to specify is μ, which was introduced in Equation (3.35). For these experiments, we defined $\mu \sim U(K_7)$ where U is the uniform distribution over a compact set.

We now aim at using Proposition 3.5.7 to approximate $\overline{V}(x)$ that corresponds to f in the proposition. Clearly $\overline{V} \in C(H, \mathbb{R})$ and K_7 is the compact subset. Let $\Pi_N : H_w \to H_w^N$ be the orthogonal projection to $H_w^N \coloneqq \text{span}\{e_1 ..., e_N\}$. The neural network parametrized by $\theta \in \mathbb{R}^M$ is defined as $f_\theta^N : H_w^N \to \mathbb{R}$ where M indicates the dimensions of the parameters. Simplifying the notation in Proposition 3.5.7, we know that for every $\delta > 0$, there exists a θ (that can depend on δ) such that for any $N \geq N^*$ the following holds true

$$\sup_{x \in K} \left| \overline{V}(x) - f_\theta^{N^\star}(x) \right| < \delta. \tag{4.2}$$

Since f_θ^N is a neural network (in [10] is a one-layer NN but the general case holds true as well), the final layer is composed of an activation function σ, in our case $\sigma : \mathbb{R} \to \mathbb{R}$. To define such function, we define a function $\sigma : H_w \to H_w^N$ of the form of

$$\sigma(x) \coloneqq \beta(\phi(x))z, \tag{4.3}$$

for $\phi \in L(H_w, \mathbb{R})$, $z \in H_w^N$ and $\beta \in Lip(\mathbb{R}, \mathbb{R})$ such that $\lim_{\xi \to \infty} \beta(\xi) = 1$, $\lim_{\xi \to -\infty} = 0$ and $\beta(0) = 0$, in particular, they specify $\beta(x) \coloneqq \max\{1 - \exp(x), 0\}$. By applying Theorem 4.3 of [9], one can note that the σ defined as such satisfies the separating property (Definition 3.5.4), moreover, being Lipschitz, it is continuous and bounded on compacts subsets; consequently, Theorem 3.5.6 holds true. To satisfy the constraint $\Im(\sigma) \subset H_w^N$, we simply choose $z \in H_w^N$ as previously indicated.

In the next Section, we will discuss how to simulate such processes and compute the corresponding option price $\overline{V}(x)$, in particular, we will discuss the case for $N = 7$.

Multi-dimensional Brownian motion

Let the 7-dimensional Brownian motion be

$$W(t) := B_1(t)e_1 + \cdots + B_7(t)e_7 \tag{4.4}$$

where B_i for $i = 1, \ldots, 7$ are one dimensional independent Brownian motions, moreover, we let $\eta = \eta^* = id$, the identity operator, in the model (3.14). According to Proposition 6.3 of [13], to ensure an arbitrage-free model, the drift becomes

$$\alpha(t, \xi) = -\frac{1}{2}\|Q^{\frac{1}{2}}\eta^*(t)(\delta\xi^*1)\|_w^2, \tag{4.5}$$

with Q being the covariance operator of the Q-Wiener process. By definition of the Brownian motion in Equation (4.4), one can see that W is a Q-Wiener process and it holds that $Q = Q^{\frac{1}{2}} = \Pi_{K_7}$ with K_7 defined in Equation (4.1), and Π_{K_7} being the orthogonal projection onto such a subset. Since $\delta_\xi^*1 = h_\xi$ (see Remark 3.5.9), the equation above simplifies to

$$\alpha(t, \xi) = -\frac{1}{2}\|\Pi_{K_7}(h_\xi)\|^2 \tag{4.6}$$

with $h_\xi(u) = 1 + \int_0^{\xi\wedge u} w^{-1}(v)dv$. Using the fact that $\Pi_{K_7}(h_\xi) = \sum_{i=1}^{7}\langle h_\xi, e_i\rangle e_i$. Furthermore, using the fact that h_ξ is the representer of the evaluation functional, in other words, $\langle h_\xi, e_i\rangle = e_i(\xi)$, we obtain

$$\Pi_{K_7}(h_\xi) = \sum_{i=1}^{7} e_i(\xi)e_i. \tag{4.7}$$

Now, simply by the definition of the norm and inner product, we obtain

$$\|\Pi_{K_7}(h_\xi)\| = \sum_{i=1}^{7} e_i(\xi)^2. \tag{4.8}$$

By plugging it in back into the Equation (4.6), we get

$$\alpha(t, \xi) = -\frac{1}{2}(e_1(\xi)^2 + \cdots + e_7(\xi)^2). \tag{4.9}$$

Putting it all together, starting from Equation (3.13), we get

$$\begin{aligned}
Y(t) &= \mathcal{S}_t Y_0 + \int_0^t \mathcal{S}_{t-s}\alpha ds + \int_0^t \mathcal{S}_{t-s}\eta dW(s) \\
&= Y_0(\cdot + t) + \int_0^t \alpha(\cdot + (t-s))ds + \int_0^t \mathcal{S}_{t-s} dW(s) \\
&= Y_0(\cdot + t) - \frac{1}{2}\int_0^t e_1(\cdot + (t-s))^2 + \cdots + e_7(\cdot + (t-s))^2 ds \\
&\quad + \sum_{i=1}^{7}\int_0^t e_i(\cdot + (t-s))dB_i(s).
\end{aligned}$$

Note that $e_i(\cdot)$ means the application of the functional e_i to its argument as ONB of a function space. For both integrals we use forward Euler-Maruyama [43] (see Appendix Section 5.3 for more information about Euler-Maruyama) where for the stochastic integral we use a sample path from the Brownian motion. Explicitly, for $i = 1, \dots, 7$ and a positive integer L, we fix a mesh $0 \le s_1 < \cdots < s_L = t$

$$\int_0^t e_i(\cdot + (t-s))^2 ds \approx \sum_{j=1}^{L} e_i(\cdot + (t-s_j))^2 (s_j - s_{j-1}). \tag{4.10}$$

In the same fashion, we apply forward Euler-Maruyama to the stochastic counterpart

$$\int_0^t e_i(\cdot + (t-s))dB_i(s) \approx \sum_{j=1}^{L} e_i(\cdot + (t-s_j))(B_i(s_j) - B_i(s_{j-1})). \tag{4.11}$$

Where $B_t(s_j)$ are sample from $N(0, \frac{t}{N})$. For the samples of the initial curve, we aim at sampling from μ which we assume to be distributed as $U([-\frac{1}{2}, \frac{1}{2}]^7)$, that is, we generate $x^{(k)} \in [-\frac{1}{2}, \frac{1}{2}]^7$ i.i.d. for each k. Once we generate $Y(t)$, we simply apply the payoff functional (3.49) to get the corresponding true label $\mathcal{X}(x^{(k)})$. Thus, the dataset will consists of the input-label pairs $(x^{(k)}, \mathcal{X}(x^{(k)}))$ for $k = 1, \dots, n$.

Explicitly, for a certain $x_0 = \sum_{i=1}^{7} \alpha_i e_i$, we have the following recursive definition:

$$x_{n+1} = x_n - \frac{1}{2}\sum_{j=1}^{L}\sum_{i=0}^{7} e_i(\cdot + (t-s_j))^2(s_j - s_{j-1}) + \sum_{i=1}^{7}\sum_{j=1}^{L} e_i(\cdot + (t-s_j))(B_i(s_j) - B_i(s_{j-1})). \tag{4.12}$$

The resulting $x_N = \widetilde{\mathcal{X}}(T, x_0)$ is the corresponding functionals with $\widetilde{\mathcal{X}}$ the Euler Maruyama approximation of $\mathcal{X}$; in our implementation, it will correspond to a time domain discretization of the function, namely the vector $(\widetilde{\mathcal{X}}(0, x_0), \dots, \widetilde{\mathcal{X}}(T, x_0)$ for $T = \frac{1}{2}$ and size L. Since we are interested in the flow forward, in accordance to Equation (3.16), we compute the integral by

$$\int_0^T \widetilde{\mathcal{X}}(T, x_0) = \frac{12}{L}\sum_{l=0}^{L} \widetilde{\mathcal{X}}(T, x_0)_l \tag{4.13}$$

We will now focus on the Neural Network component.

Dataset generation

First, we start by generating the training and test dataset. To replicate the results of the original paper, we generate the same amount of data. The training dataset contains $n = 10,000,000$ samples and is obtained by generating the datapoints of size M which are represented as vectors, i.e. $x^{(k)} = (x_1^{(k)}, \dots, x_M^{(k)}) \in \left[-\frac{1}{2}, \frac{1}{2}\right]^M$. Note that although this is a M-dimensional vector, it actually represents the projection to the ONB basis outlined in Equation (5.33). The corresponding label is then computed using the model (3.13) and the corresponding option price at Equation (3.48) using the approximations outlined in the previous paragraph for both the Lebesgue integral and the stochastic integral using a realization of the Brownian motion, thus the output pairs will be of the form $(x^{(k)}, \mathcal{X}(x^{(k)}))$ for $k = 1, \dots, n$.

Conversely, the test data is slightly different; in fact, the label is $\mathbb{E}\left[\mathcal{X}(x^{(k)})\right]$ instead of simply $\mathcal{X}(x^{(k)})$. To compute this, we used forward Euler Monte-Carlo with the Monte-Carlo simulations being $n_{MC} = 100,000$. Moreover, the dataset size is $n = 10,000$ in this case, as the test dataset needs not to be as large as the training one. Finally, for the test set, we used a fixed projection size of 10.

We generate the dataset using Pytorch on CUDA to leverage the GPU and compute the price in batches; we were able to generate the whole dataset is around 10 seconds, compared to the original paper that requires 43 hours. The discrepancy is due to implementation: in contrast to the original work, where the dataset was generated by discretizing and simulating the SPDE on CPU (which is computationally very demanding, especially in the multi-dimensional case), we used PyTorch with CUDA and batch computations. This allows millions of samples to be simulated in parallel on GPU, which is more suited for this task, and since we only require the projection onto 7 basis functions, the dataset generation reduces to simple vectorized operations, which explains why it only takes about 10 seconds instead of several hours.

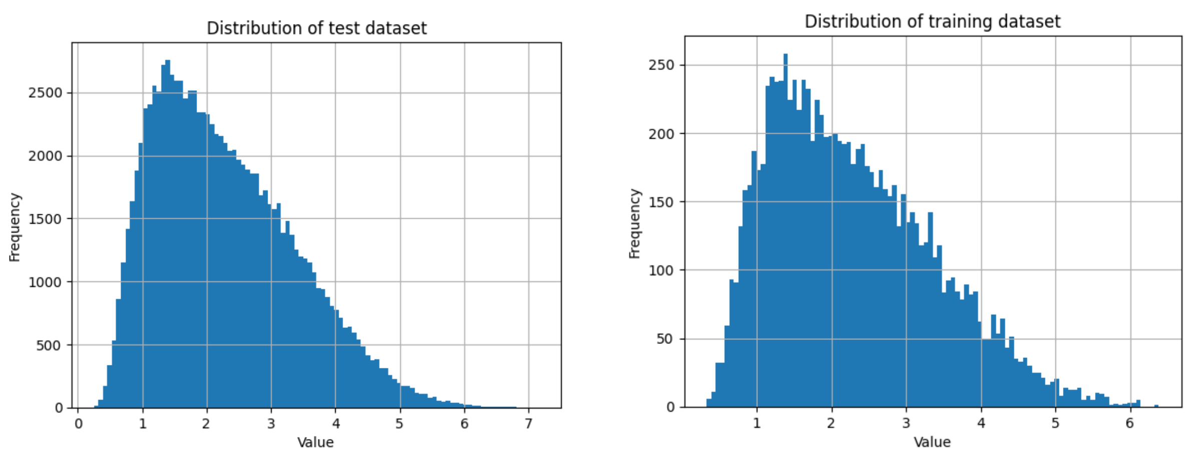

Figure 4.1: The distributions of the price functional $\mathcal{X}(x)$ for the test and train datasets.

Training of the Neural Network

We ran the following procedure, using $M = 2, \dots, 10$, with both basis as in Equation

(5.33) as well as the Laguerre functional basis describe in Equation (5.32). We trained two different models of NN, one composed of linear combination of the single neuron neural network described by Equation (2.7). In particular, such network is composed of several of the so-called "paths" that are computed in parallel, whose outputs are then summed at the end. Each of these paths is composed of a dense layer with output dimensions of 7, followed by the custom activation function described in Equation (4.3), and finally followed by a dense layer with output dimensions of 1. The second model we train is a vanilla deep neural network with several 5 linear layers alternating with ReLU ($\mathrm{ReLU}(x) = \max(0, x)$) activation function.

We train the models using the standard ADAM optimizer [41] and using L^2 distance (Mean Squared error) as the loss function.

We noted that adding weight decay of 10^{-3} significantly improved the convergence rate during training for both models. We performed SGD with a batch size of 10000 exactly like the original paper. We run the training for 100 epochs. Both models successfully converged, however, we did not notice any substantial difference between the multi-paths and the standard MLP model, with the former one achieving a test mean squared error of 0.006 and the standard model of 0.0003. On the contrary, the convergence speed was in all the tests sensitively faster for the multi-layer base network. An important reason for this might not be worse architecture, but merely due to the number of trainable parameters. We report below the converge of both training and test for different M for both the multi-paths and multi-layer networks.

Noise dimension	Multipath Train Loss	Multipath Val Loss	Standard Train Loss	Standard Val Loss	Multipath Trainable Params	Standard Trainable Params
2	0.004508	0.060314	0.001343	0.057136	80	2945
3	0.003685	0.023585	0.001495	0.021335	150	3009
4	0.002977	0.008296	0.001396	0.006663	240	3073
5	0.003566	0.004190	0.001484	0.002086	350	3137
6	0.003790	0.003218	0.001523	0.001038	480	3201
7	0.004457	0.003739	0.001463	0.000888	630	3265
8	0.004365	0.003537	0.001438	0.000819	800	3329
9	0.012238	0.009464	0.001517	0.000571	990	3393
10	0.006923	0.005362	0.001566	0.000276	1200	3457

Table 4.1: Convergence result using the Laguerre functionals as basis (Equation (5.32)).

Our full code is available on GitHub. `https://github.com/AlviseSembenico/Flow-forward-SPDE`

Noise dimension	Multipath Train Loss	Multipath Val Loss	Standard Train Loss	Standard Val Loss	Multipath Trainable Params	Standard Trainable Params
2	0.025485	0.155383	0.003509	0.135788	80	2945
3	0.010627	0.056276	0.003763	0.048758	150	3009
4	0.012577	0.018419	0.003962	0.008912	240	3073
5	0.010926	0.007184	0.004092	0.000671	350	3137
6	0.010086	0.006949	0.003896	0.000484	480	3201
7	0.019343	0.016247	0.004076	0.000507	630	3265
8	0.012317	0.010508	0.003789	0.000336	800	3329
9	0.016278	0.012612	0.004003	0.000349	990	3393
10	0.012991	0.008888	0.003723	0.000304	1200	3457

Table 4.2: Convergence result using the basis in Equation (5.33).

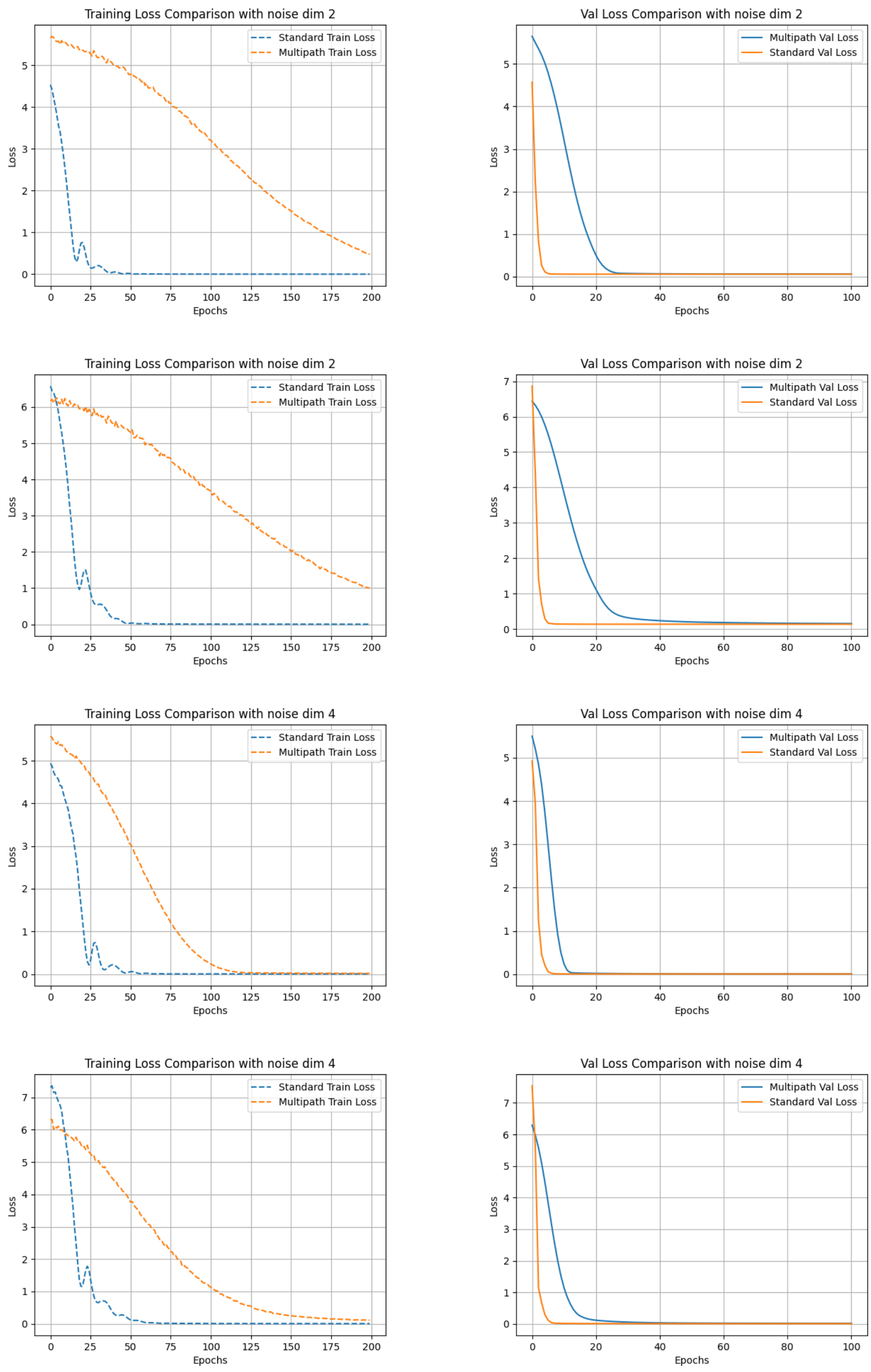

Figure 4.2: Plot of training and validation loss of the two neural networks with The Laguerre and polynomial basis with noise dimension of 2 and 4.

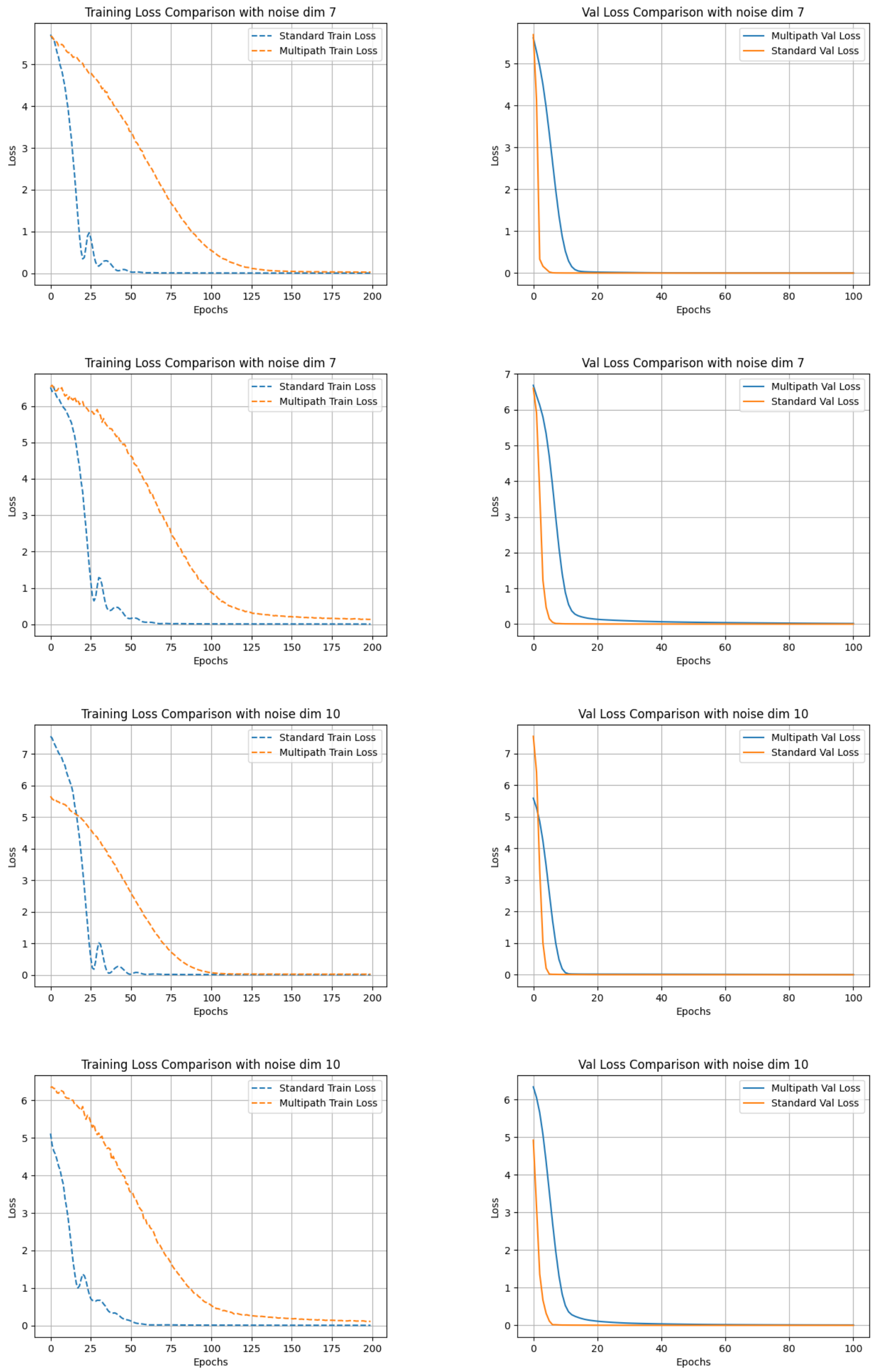

Figure 4.3: Plot of training and validation loss of the two neural networks with The Laguerre and polynomial basis with noise dimension of 7 and 10.

5 Appendix

5.1 Topological space

We recall the concept of seminorm which is used as in the density argument for the neural network approximation of the forward price functional.

Definition 5.1.1. *For a vector space X over a field F, a function $p : X \to \mathbb{R}$ is a seminorm if the following conditions hold.*

- *Subadditivity: $p(x + y) \leq p(x) + p(y)$.*
- *Homogeneity: $p(sx) = |s|p(x)$ for $s \in F$.*

5.2 Stochastic Integration

We recapitulate the most important definitions from standard Stochastic Analysis that we will directly make use of. For a full account on the subject, you can consult [55].

Definition 5.2.1. *Given a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\leq T}, P)$, the filtration $\{\mathcal{F}_t\}_{t\leq T}$ is said to be right continuous if*

$$\mathcal{F}_t = \bigcap_{s\geq t} \mathcal{F}_s. \tag{5.1}$$

Definition 5.2.2. *A filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\leq T}, P)$ is said to satisfy the usual conditions if $\{\mathcal{F}_t\}_{t\leq T}$ is right continuous, and $\mathcal{F}_0$ contains the null set, i.e.*

$$\{\omega \subset \Omega | \exists F \in \mathcal{F} : \omega \subset F \text{ and } P(F) = 0\} \subset \mathcal{F}_0. \tag{5.2}$$

Definition 5.2.3. *Let X a space. A family A of functions $f : X \to \mathbb{R}$ is called point separating if for every two distinct points $x, y \in X$, there exists a function $f \in A$ such that $f(x) \neq f(y)$.*

Definition 5.2.4. *Let X a space. A family A of functions $f : X \to \mathbb{R}$ is called vanishing nowhere if for every $x \in X$ there exists some $f \in A$ with $f(x) \neq 0$.*

5.2.1 Measures

We now examine some measures and processes used in finance to model jumps. In particular, the Poisson process is used to model the jump amplitude, while the Lévy measure defines the distribution of the occurrence of the jumps.

Lévy measure and Stochastic integration
We report the Definition II.1.20 of [38] regarding the Poisson measure.

Definition 5.2.5. *A Poisson measure on $\mathbb{R}^+ \times H$ relative to the filtration $\mathbb{F}$, is an integer-valued measure μ such that*

- *the measure m on $\mathbb{R}^+ \times H$ defined by $m(A) = \mathbb{E}[\mu(A)]$ is σ-finite.*
- *for every $s \geq 0$ the variable $\mu(\cdot, A)$, where $A \subset (s, \infty) \times H$ is independent of the σ-field $\mathcal{F}_s$.*

If m is such that $m(\{t\} \times H) = 0$ for all $t \geq 0$, then μ is called Poisson Measure The measure m is called intensity measure

Definition 5.2.6. *Let m be defined as in Definition 5.2.5. If m has the form $m(dt, dx) = dt \times F(dx)$ (is the product measure) for F a positive σ-finite measure on $(H, B(H))$, then μ is called homogeneous Poisson measure. We call $dt \times F(dx)$ the compensator of μ.*

One can think of μ as the random measure that determines how many jumps are in a set $A \subset \mathbb{R}^+ \times H$, while m is by definition the expected mass of μ in a set A. Note that μ is defined as $\mu : (\Omega \times \mathbb{R}^+ \times H)$, which implies that $\mu(B)$ for $B \subset \mathbb{R}^+ \times H$ is again a stochastic process.

Definition 5.2.7. *A Borel measure ν on $\mathbb{R}$ is a Lévy measure if $\nu(0) = 0$ and*

$$\int_{\mathbb{R}} (1 \wedge |x^2|)\nu(dx) < \infty. \tag{5.3}$$

Although this definition might seem arbitrary, the integrability condition comes from the Lévy–Khintchine representation Theorem, which gives an explicit formula for the characteristic function of a general Lévy process; however, this is beyond the scope of this work. In practice, for a set A, the measure $\nu(A)$ quantifies the expected number of jumps per unit of time whose size falls into the set A.

We refer to [31] and [52] for an extensive analysis of the subject, however, for the sake of completeness, in this work we briefly introduce the concept of stochastic integration with respect to a homogeneous Poisson measure with compensator following section 2.3 of [27].

We define the Itô integral for the compensator, that is $\mu(ds, dx) - F(dx)ds$. For a suitable integrable function Φ, the Itô integral is the isometry

$$\mathbb{E}\left[\left\|\int_0^t \int_E \Phi(s,x)(\mu(ds,dx) - F(dx)\,ds)\right\|^2\right] = \mathbb{E}\left[\int_0^t \int_E \|\Phi(s,x)\|^2 F(dx)\,ds\right]. \tag{5.4}$$

In practice, this defines the stochastic integration over a Hilbert-space-valued Lévy measure, which is used to model jumps in such settings.

For a Borel measurable set $B \subset H$ with $m([0,\infty), B) < \infty$, and the Poisson process $N_t := \mu((0,t] \times B)$, if we defined $\lambda_B := m([0,1] \times B)$, we have $\mathbb{E}[N_t(B)|\mathcal{F}_s] = N_s(B)$ and $\mathbb{E}[N_t(B)] = \lambda_B t$ from direct computation. Therefore, $N_s(B)$ is clearly not a martingale. To turn it into a martingale, we use the concept of a compensator, which has been introduced in the definition above. Thus, we can speak of $\widetilde{\mu}(dt, dx) := \mu(dt, dx) - m(dt, dx)$ as a squared integrable martingale, which makes the Equation (5.4) well defined.

We now state the Vitali converge theorem, which gives equivalent conditions for a sequence of random variables that converges in L^p norm, to converge in probability.

Theorem 5.2.8 (Theorem 4.5.4 from [17]). *Let μ be a finite measure. Suppose that f is a μ-measurable function and $\{f_n\}$ is a sequence of μ-integrable functions. Then, the following assertions are equivalent:*

(i) the sequence $\{f_n\}$ converges to f in measure and is uniformly integrable;

(ii) the function f is integrable and the sequence $\{f_n\}$ converges to f in the space $L^1(\mu)$.

5.3 Euler-Maruyama

In this section, we outline the most widely used numerical method for finite-dimensional stochastic differential equations (SDEs), the *forward Euler–Maruyama scheme.* Let $X = (X(t))_{t\in[0,T]}$ be an $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$-adapted process on the probability space $(\Omega, \mathcal{F}, P)$, satisfying the SDE

$$\begin{cases} \dot{X}(t) = a\big(t, X(t)\big)\,dt \;+\; b\big(t, X(t)\big)\,dW_t, \quad 0 \leq t \leq T, \\ X(0) = x_0, \end{cases} \tag{5.5}$$

where $W = (W_t)_{t\geq 0}$ is a standard Brownian motion and $a, b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ are globally Lipschitz in the spatial variable and measurable in t. A (strong) solution is characterized by

$$\mathbb{E}\left[\left(X(t) - \left(x_0 + \int_0^t a(s, X(s))\,ds + \int_0^t b(s, X(s))\,dW_s\right)\right)^2\right] = 0 \quad \text{for all } t \in [0,T].. \tag{5.6}$$

Fix a partition $\pi : \; 0 = t_0 < \cdots < t_N = T$ of $[0,T]$ with mesh size $|\pi| := \max_{0\leq n\leq N-1}(t_{n+1} - t_n)$. Set $\Delta t_n := t_{n+1} - t_n$ and $\Delta W_n := W_{t_{n+1}} - W_{t_n}$. The discrete Euler–Maruyama approximation $\widetilde{X} = (\widetilde{X}(t_n))_{n=0}^N$ is defined recursively by

$$\begin{cases} \widetilde{X}(t_{n+1}) := \widetilde{X}(t_n) + a\big(t_n, \widetilde{X}(t_n)\big)\,\Delta t_n + b\big(t_n, \widetilde{X}(t_n)\big)\,\Delta W_n, \\ \quad \widetilde{X}(0) := x_0. \end{cases} \tag{5.7}$$

Furthermore, we can extend the discrete version we just defined to all $t \in [0,T]$ as follows

$$\widetilde{X}(t) := \widetilde{X}(t_n) + \int_{t_n}^{t_{n+1}} a(t_n, \widetilde{X}(t_n))ds + \int_{t_n}^{t_{n+1}} b(t_n, \widetilde{X}(t_n))dW_s, \tag{5.8}$$

where the integral $\int_{t_n}^{t_{n+1}} a(t_n, \widetilde{X}(t_n))ds$ is Lebesgue-Stieltjes (see Appendix C of [55]) and the latter one is standard Itô-integral.

The main result concerning this approximation is the following, which we state without proof; consult [18] for a comprehensive analysis of the subject.

Theorem 5.3.1. *Let $\widetilde{X}_{\Delta t}(t)$ the approximation of $X(t)$ defined as in Equation (5.8) with $|\pi| = \Delta t$, and a, b being jointly Lipschitz, then there exists a constant K_T not depending on Δt such that*

$$\max_{t\in[0,T]} \mathbb{E}\left[\widetilde{X}(t)^2\right] \leq K_T, \tag{5.9}$$

and

$$\lim_{N\to\infty} \mathbb{E}\left[|\widetilde{X}_{\Delta t^N}(t) - X(t)|\right] = 0, \tag{5.10}$$

with $\Delta t^N \to 0$.

In other words, we have strong convergence of the forward Euler Maruyama scheme.

To simulate such scheme in practice, we consider $N := \frac{1}{\Delta t}$ and $M \in \mathbb{N} < \infty$. We then run Monte Carlo to generate M independent paths, one for each ω_i with $i = 1, \ldots, M$ using Equation (5.6) and then compute the mean which, by the Central Limit Theorem, will converge to $\mathbb{E}[X(T)]$ weakly. Formally, we compute

$$\mathbb{E}\left[\widetilde{X}(T)\right] = \frac{1}{M}\sum_{i=1}^{M} \widetilde{X}(T)(\omega_i), \tag{5.11}$$

with

$$\widetilde{X}(T) = x_0 + \sum_{n=1}^{N} a(t_n, X(t_n))\Delta_{t_n} + \sum_{n=1}^{N} b(t_n, X(t_n))W_{t_n}. \tag{5.12}$$

5.4 Calculus in Hilbert space

In the section on the Stochastic integral in separable Hilbert space, as well as in the definition of SPDEs, the drift term is defined in terms of Bochner integral. Moreover, for the Itô's lemma that we extended from the real case, the concept of Fréchet derivative is of crucial importance. The primary purpose of this section is to provide the essentials to properly understand these objects. The following setting will be slightly more general, involving functions from Ω to B where B is an arbitrary Banach space. We follow [1] for this section with the needed adaptations.

5.4.1 Bochner Integral

We start by defining the concept of Hilbert space-valued simple functions.

Definition 5.4.1. *Let $\Omega, \mathcal{F}, \mu$ be a measure space. The set of simple functions is defined as follows:*

$$\mathcal{E} := \left\{ f : \Omega \to H : f = \sum_{k=1}^{m} h_k \mathbb{1}_{A_k} \quad \text{with } h_k \in H, A_k \in \mathcal{F}, k \in \mathbb{N} \right\}. \tag{5.13}$$

Moreover, we endow $\mathcal{E}$ with the seminorm

$$\| \cdot \|_{\mathcal{E}} := \int_H \| \cdot \| d\mu. \tag{5.14}$$

Note that the seminorm defined above is not a norm, as the different functions f, g can differ on a measure-zero set. In the same fashion we do for Lebesgue integral, we consider equivalence classes of functions, namely we consider $f \sim g$ if $f = g$, μ a.s. Note that we keep the same syntax, and when we consider a function f, we are de facto considering a representative of such equivalence class. It direct to see that all the following definitions are indeed representative independent, thus well defined. We now consider each $f \in \mathcal{E}$, i.e. each $A_k \cap A_j$ for $j \neq k$. One can easily prove that this representation is always possible. We then defined the Bochner integral as follows.

Definition 5.4.2. *For any $f \in \mathcal{E}$ we define the Bochner integral as*

$$\int f d\mu := \sum_{k=1}^{n} h_k \mu(A_k). \tag{5.15}$$

The idea to extend this definition is standard work, namely, consider the function class in which $\mathcal{E}$ is dense, and then take the linear extension of the integral. We will give a brief construction of $\overline{\mathcal{E}}$.

Definition 5.4.3. *A function $f : \Omega \to H$ is called strongly measurable if it is $\mathcal{F}/B(H)$-measurable and $F(\Omega) \subset H$ is separable.*

We now define the $\mathcal{L}^p$ space as follows.

Definition 5.4.4. *tLet*

$$\mathcal{L}^p(\Omega, \mathcal{F}, \mu, H) := \{ f : \Omega \to H : f \text{ is strongly measurable w.r.t. } \mathcal{F}, \\ \text{and } \int_H \|f\|^p d\mu < \infty \}. \tag{5.16}$$

The corresponding semi-norm is

$$\|f\|_{L^p} := \left(\int_H \|f\|^p d\mu \right)^{\frac{1}{p}}. \tag{5.17}$$

where $L^p(\mu, H)$ is the space of equivalence classes of $\mathcal{L}^p(\Omega, \mathcal{F}, \mu, H)$ with respect to $\| \cdot \|_{L^p}$.

We now state and prove the lemma that will help us to prove the final result.

Lemma 5.4.5 (Lemma A.1.4 from [1]). *Let H be a metric space with metric d, let $f : \Omega \to H$ strongly measurable. Then, there exists a sequence f_n of simple H-valued functions such that $\sup_\omega d(f_n(\omega), f(\omega))$ converges monotonely to zero.*

Proof. Take $\{e_i\}_{i\in\mathbb{N}}$ a dense subset of $f(\Omega)$. Note that such a dense subset exists by the assumption that f is strongly measurable, which implies that $f(\Omega)$ is separable. Let $\omega \in \Omega$ be arbitrary. For $m \in \mathbb{N}$, define

$$\begin{aligned} d_m(\omega) &:= \min\{\, d(f(\omega), e_k) \mid k \leq m \,\} \\ k_m(\omega) &:= \min\{\, k \leq m \mid d_m(\omega) = d(f(\omega), e_k)\}, \\ f_m(\omega) &:= e_{k_m(\omega)}. \end{aligned}$$

Note that f_m are simple functions and measurable, moreover $f_m(\Omega) \in \{e_1, \dots, e_m\}$ by definition. By the density of $\{e_i\}_{i\in\mathbb{N}}$, we have that $d_m(\omega)$ is monotonically decreasing to 0 for any $\omega \in \Omega$, the assertion follows. $\square$

Thus, for any $f \in L^1$, by the Lemma just proven, there exists an approximating sequence $f_n \to f$ in the supremum topology. By the Lebesgue dominated convergence theorem, we have that $f_n \to f$ in L^1 sense. We omit the remaining proof to show that L^1 is complete and closed under pointwise limit. These facts, together with the just prove lemma shows that $\mathcal{E} = L^1$, and the corresponding integral of an arbitrary $f \in L^1$ is defined as $\lim_{n\to\infty} \int_h f_n d\mu$ for an approximating sequence f_n.

It is important to mention that most of the important property that holds for Lebesgue integral hold for the Bochner integral as well such as the triangle inequality, and the Fundamental theorem of Calculus in case the integrating function is of the type $f \in C^1([a, b], H)$; therefore you can expect to manipulate the Bochner integral in the same fashion for what concerns the content of this work.

5.5 Doob maximal inequalities

Proposition 5.5.1 (Proposition 2.2 of [55]). *Let $M_t \in L^p(\Omega, \mathcal{F}, P)$ is an H-valued submartingale, then the following inequalities hold true*

$$P\left(\sup_{0\leq t\leq T} \|M_t\|_H > \lambda\right) \leq \frac{1}{\lambda^p}\mathbb{E}\|M_T\|_H^p, \quad p \geq 1,\ \lambda > 0; \tag{5.18}$$

$$\mathbb{E}\left(\sup_{0\leq t\leq T} \|M_t\|_H^p\right) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}\|M_T\|_H^p, \quad p > 1. \tag{5.19}$$

For the proof, consult the reference.

5.6 Forwards

In this section, we give a brief overview of forward and option pricing. Applications thereof will be discussed in Chapter 3.2 regarding SPDE and Chapter 3.5 on the topic of forward option pricing. For a complete account on the subject, consult [13].

Commodities such as crude oil, agricultural products, and energy are traded over-the-counter on exchanges worldwide. The financial products are split into spot price, referring to immediate delivery, and forward and futures, referring to delivery in the future, as well as all derivative products, such as options (European calls and puts are the most liquid).

The spot price refers to the closest future or forward contract with the closest delivery period, that is, the closest delivery time in the future. In fact, for the electricity market, the spot price refers to the day-ahead market, namely the delivery for the same hour of the next day, so there is always a 24-hour difference between the last tradable moment of the future and the actual delivery time. This is by design so that energy producers have the physical time to increase/decrease the energy output based on the demand, since energy storage can still be considered practically not feasible on any non-trivial scale.

Future and forward can have both a physical delivery or be used as a speculation tool, where the buyer receives the spot price value in money rather than the commodity delivery itself. However, this distinction, as well as the difference between forward and future, does not play a role in this work; therefore, we will refer to the general class simply as futures.

Negative pricing and high volatility should be taken into account when mathematically modeling forward prices. Contrary to the equity market, such as the stocks and bonds market, when oversupplied, the energy market can show negative prices due to the costs of reducing energy production. Moreover, jumps due to events, seasonality, and weather can significantly affect the price.

Stochastic modeling

In the same fashion as Chapter 3.4 for $S(t)$ modeling the spot price at time $t \geq 0$ and which is a local martingale, the forward price for delivery at time T can be written as

$$F(t,T) = \mathbb{E}_{\mathbb{Q}}\left[S(T)|\mathcal{F}_t\right]. \tag{5.20}$$

where $\mathcal{F}_t$ encapsulates the information known to the market at time t and the measure $\mathbb{Q}$ is known as the martingale measure or risk-neutral.

In the most simple form, $F(t,\cdot)$ can be modeled as a Stochastic differential equation

$$dF(t,T) = \alpha(t,T)\,dt + \sigma(t,T)\,dB(t). \tag{5.21}$$

It is preferable to rephrase the problem as time to delivery rather than time of delivery using the reparameterization $x \coloneqq T - t$, and defining the corresponding functional

$$f(t,x) = F(t,x+t). \tag{5.22}$$

We can then recast once again the problem, looking at the object $f(t,\cdot)$ which is a functional in an appropriate function space, which is generally chosen to be a separable Hilbert space. As the study [44] suggests, having several driving Brownian motions to model can be cumbersome, which suggests choosing an infinite-dimensional noise as the natural choice. Therefore, the price functional $f(t,\cdot)$ is generally modeled as an SPDE. In reparameterization $x \coloneqq T-t$, the no-arbitrage dynamics of f will have the differential operator $\frac{\partial f(t,x)}{\partial x}$. Since the differential operator ∂f is not a bounded operator, we opt to model it as a family of C_0 semigroups. Following the Chapters 1.3.1 and 1.1, it becomes natural to work with mild solutions of such SPDE.

Given a separable Hilbert space H, we can model $f(t,x)$ via the following SPDE:

$$df(t) = \partial_x f(t)dt + \beta(t)dt + \phi(t)d\widetilde{N}(t), \tag{5.23}$$

with $\widetilde{N}$ being the H-valued martingale compensated version of the Lévy process, which has been defined in Section 5.2.1, and $f(0) = f_0 \in H$. For the path-dependent version, we can simply replace the SPDE above with

$$df(t) = \partial_x f(t)dt + \beta(t,f(t))dt + \psi(t,f(t))d\widetilde{N}(t), \tag{5.24}$$

which is equivalent to the formulation of the SPDE in Equation (3.13). In the work of Filipović [27], the existence and uniqueness under certain conditions of the coefficients are shown.

To conclude this section, we present the result that motivates us to use the Filipović space introduced in Definition 3.1.3. We first recall its definition.

Definition 5.6.1. *Define H_w as the space of real-value, absolutely continuous functions on $\mathbb{R}^+$ satisfying the norm*

$$|f|_w^2 \coloneqq f(0)^2 + \int_0^\infty w(x)f'(x)^2 dx, \tag{5.25}$$

with f' being the weak derivative of f; w is an increasing function with $w(0) = 1$ and inverse integrable, i.e. $\int_{\mathbb{R}^+} w^{-1}(x)dx < \infty$

Lemma 4.2.1 in [25] helps us understand why the choice of the Filipović space is useful.

Proposition 5.6.2 (Theorem 6.10 from [28])**.** *H_w is a separable Hilbert space with ∂_x is the densely defined generator of the C_0 semigroup $S(t)g = g(\cdot + t)$. Moreover, the semigroup is pseudo-contractive and uniformly bounded, i.e.*

$$\|S(t)\|_{\text{op}} \leq e^{kt} \quad \|S(t)\|_{\text{op}} \leq K \tag{5.26}$$

for k, K positive constants.

The final proposition that gives us the tool to work in the Filipović space is the following from [8].

Proposition 5.6.3. *After a proper rescaling of the norm $|\cdot|_w$, the Filipović space H_w is a Banach algebra.*

Proof. Define $k^2 \coloneqq \int_0^\infty w^{-1}(x)dx$ such that $k < \infty$. Let

$$h_x(y) \coloneqq 1 + \int_0^{x \wedge y} w^{-1}(z)dz. \tag{5.27}$$

Therefore, for a $g \in H_w$, we have $g(x) = \delta_x g$. Note that $\langle g, h_x \rangle_w = g(x)$ by the following, by the definition of the Filipović space norm in Definition 3.1.3.

$$\langle g, h_x \rangle_w \coloneqq g(0)h_x(0) + \int_0^\infty w(z)g'(z)h_x(z)dz$$

Note that $h_x(0) = 1$ and $h'_x(z) = w^{-1}(z)\mathbb{1}_{[0,x\wedge y]}$; by putting these equalities back into the equation above, we get:

$$\langle g, h_x \rangle_w =: g(0) + \int_0^x w(z)g'(z)w^{-1}(z)dz = g(0) + [g'(z)]_0^x = g(x). \tag{5.28}$$

Going back to $g(x) = \delta_x g(x)$, we can see that $\delta_x g = \langle h_x, g \rangle_w$. Thus, we have

$$|g(x)|^2 \leq |\langle h_x, g \rangle_w|^2 \leq |h_x|_w^2 |g|_w^2 \tag{5.29}$$

by Cauchy-Schwarz. Moreover,

$$|h_x|_w^2 |g|_w^2 = h_x(x)|g|_w^2 \leq (1 + k^2)|g|_w^2. \tag{5.30}$$

Applying the above and using the product rule for the derivative, we arrive at the following:

$$|fg|_w^2 \leq (t + 4k^2)|f|_w^2 |g|_w^2, \tag{5.31}$$

which concludes the proof. □

The last two results give us the tools necessary to work confidently in such a space, leveraging the properties of Banach algebra to properly define the semigroup generator in the exponential form and being closed under multiplication. The application of such space is analyzed in full detail in Section 3.5.

5.7 Appendix

5.7.1 Laguerre functional basis

Here we write the image of the first 10 Laguerre functional through J as described in Section 3.5.1.

$$
\begin{aligned}
\Phi_0(\xi) &= 1, \\
\Phi_1(\xi) &= e^{-\xi}, \\
\Phi_2(\xi) &= \xi e^{-\xi}, \\
\Phi_3(\xi) &= \left(\frac{\xi(2-\xi)}{2}\right) e^{-\xi}, \\
\Phi_4(\xi) &= \left(\frac{\xi(\xi^2-6\xi+6)}{6}\right) e^{-\xi}, \\
\Phi_5(\xi) &= \left(\frac{\xi(-\xi^3+12\xi^2-36\xi+24)}{24}\right) e^{-\xi}, \\
\Phi_6(\xi) &= \left(\frac{\xi(\xi^4-20\xi^3+120\xi^2-240\xi+120)}{120}\right) e^{-\xi}, \\
\Phi_7(\xi) &= \left(\frac{\xi(-\xi^5+30\xi^4-300\xi^3+1200\xi^2-1800\xi+720)}{720}\right) e^{-\xi}, \\
\Phi_8(\xi) &= \left(\frac{\xi(\xi^6-42\xi^5+630\xi^4-4200\xi^3+12600\xi^2-15120\xi+5040)}{5040}\right) e^{-\xi}, \\
\Phi_9(\xi) &= \left(\frac{\xi(-\xi^7+56\xi^6-1176\xi^5+11760\xi^4}{40320}\right. \\
&\qquad \left. + \frac{\xi(-58800\xi^3+141120\xi^2-141120\xi+40320)}{40320}\right) e^{-\xi},
\end{aligned}
\tag{5.32}
$$

5.7.2 Exponential function basis

$$
\begin{aligned}
e_1(\xi) &= 1,\\
e_2(\xi) &= \exp(-\xi) - 1,\\
e_3(\xi) &= \xi \exp(-\xi),\\
e_4(\xi) &= \frac{1}{2}\left(\xi^2 - 2\xi\right) e^{-\xi},\\
e_5(\xi) &= \frac{\xi^3 - 36\xi^2 - 6\xi}{42\sqrt{5}} e^{-\xi},\\
e_6(\xi) &= \frac{\xi^4 - 1440\xi^3 - 192\xi^2 - 24\xi}{24\sqrt{806115}} e^{-\xi},\\
e_7(\xi) &= \frac{\xi^5 - 100800\xi^4 - 10800\xi^3 - 1200\xi^2 - 120\xi}{1560\sqrt{49407661}} e^{-\xi}.\\
e_8(\xi) &= \frac{\xi^6 - 30\xi^5 + 300\xi^4 - 1200\xi^3 + 1800\xi^2 - 720\xi}{720}\, e^{-\xi},\\
e_9(\xi) &= \frac{\xi^7 - 42\xi^6 + 630\xi^5 - 4200\xi^4 + 12600\xi^3 - 15120\xi^2 + 5040\xi}{5040}\, e^{-\xi},\\
e_{10}(\xi) &= \frac{\xi^8 - 56\xi^7 + 1176\xi^6 - 11760\xi^5 + 58800\xi^4 - 141120\xi^3 + 141120\xi^2 - 40320\xi}{40320}\, e^{-\xi}.
\end{aligned}
\tag{5.33}
$$

Popular summary

When you buy or sell electricity or gas for delivery in the future, you are not trading a single number, you are trading along a *curve* of future delivery times. Markets therefore quote *forward contracts*, whose prices say, for each future date T, what one unit of energy for delivery at T costs today. Many contracts deliver over an entire period (say, next month), not just a single day; these are called *flow forwards*. Options on such products pay off depending on the *average* forward price over the delivery window.

Because the whole *forward curve* moves randomly through time, a natural mathematical model treats it as a time-evolving function. In this thesis we describe that evolution with a *stochastic partial differential equation* (SPDE). You can think of an SPDE as a "noisy" version of a PDE: instead of tracking one number, we track a function of the delivery time, and randomness enters through a generalised multi-dimensional noise.

Numerically working with SPDEs is hard. If we discretise the curve at many delivery times, we obtain a very high-dimensional system. Standard Monte Carlo methods (simulate many scenarios and average) become expensive: to keep the error small, the number of scenarios typically grows extremely fast with the dimension. This is the famous *curse of dimensionality*.

A different idea is to approximate the desired answer—here, the option price—as a function of the current forward curve using a *neural network*. A neural network is a flexible formula built from many simple "neurons." There is a classical mathematical result, the *universal approximation theorem*, which says that under mild conditions these networks can approximate any continuous function as closely as we like (on suitable sets). This gives a path to efficient algorithms even in large or infinite dimensions: if we can show the option price depends *continuously* on the current curve, then a neural network should be able to learn it.

The thesis makes this precise in three steps. First, we write the option price as a *functional* of today's forward curve—concretely, as the expected payoff of the average of future forward prices over the delivery period. Second, we prove this price functional is *Lipschitz*: small changes in the current curve lead to proportionally small changes in the price. This is the key continuity property. Third, we place neural networks directly on the function space where the forward curve lives (a Hilbert space of functions that supports "shifting" the curve and evaluating it at a point). Under a natural condition on the activation function (it must be sufficiently "separating"), we obtain a universal approximation result: finite neural networks can approximate the price functional arbitrarily well on compact sets.

Real markets are not compact, so we also discuss approaches that handle the whole space. Two complementary frameworks are considered: *weighted spaces*, which control growth at infinity and allow for global approximation theorems, and a finite-dimensional result for activations that behave *linearly at infinity*, which yields uniform approximation on all of $\mathbb{R}^n$. We compare assumptions and implications of these methods for models built on forward curves.

Finally, we test the theory numerically. We project the curve onto finitely many features (so the network remains finite), train the network to match simulated option values, and observe that the learned networks reach competitive accuracy with fewer simulations than standard Monte Carlo would need. In short: by combining a careful infinite-dimensional model with the approximation power of neural networks, we obtain a practical, mathematically justified tool for pricing options on flow forwards.

Bibliography

[1] *A concise course on stochastic partial differential equations.* Vol. 1905. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2007.

[2] P. Achlioptas. *Stochastic gradient descent in theory and practice.* PhD thesis, Stanford University. 2019.

[3] N. I. Akhiezer. *The classical moment problem and some related questions in analysis.* Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 2020.

[4] L. Ambrosio, G. Da Prato, and A. Mennucci. *Introduction to measure theory and integration.* Edizioni della Normale, 2011.

[5] K. Andersson, A. Andersson, and C. W. Oosterlee. "Convergence of a robust deep FBSDE method for stochastic control". In: *SIAM Journal on Scientific Computing* 45 (2023), A226–A255.

[6] A. Barth and A. Lang. "Simulation of stochastic partial differential equations using finite element methods". In: *Stochastics: An International Journal of Probability and Stochastic Processes* 84.2-3 (2012), pp. 217–231.

[7] C. Beck, W. E, and A. Jentzen. "Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations". In: *Journal of Nonlinear Science* 29 (2019), pp. 1563–1619.

[8] F. E. Benth. *Lecture II – Analysis of the forward price dynamics – infinite dimensional approach.* Lecture notes.

[9] F. E. Benth, N. Detering, and L. Galimberti. "Neural networks in Fréchet spaces". In: *Annals of Mathematics and Artificial Intelligence* 91 (2023), pp. 75–103.

[10] F. E. Benth, N. Detering, and L. Galimberti. *Pricing options on flow forwards by neural networks in Hilbert space.* arXiv preprint. 2022.

[11] F. E. Benth, N. Detering, and P. Kruehner. *Stochastic Volterra integral equations and a class of first order stochastic partial differential equations.* arXiv preprint. 2020.

[12] F. E. Benth and S. Koekebakker. "Stochastic modeling of financial electricity contracts". In: *Energy Economics* 30.3 (2008), pp. 1116–1157.

[13] F. E. Benth and P. Krühner. *Stochastic models for prices dynamics in energy and commodity markets: an infinite-dimensional perspective.* Springer, 2023.

[14] K. Bhattacharya et al. “Model reduction and neural networks for parametric PDEs”. In: *SMAI Journal of Computational Mathematics* 7 (2021), pp. 121–157.

[15] T. Björk. *Arbitrage theory in continuous time*. Oxford: Oxford University Press, 2004.

[16] P. Blanchard and E. Brüning. “Hilbert–Schmidt and trace class operators”. In: *Mathematical Methods in Physics* (2015), pp. 365–391.

[17] V. I. Bogachev. *Measure theory*. Springer, 2007.

[18] K. Burrage, P. Burrage, and T. Mitsui. “Numerical solutions of stochastic differential equations – implementation and stability issues”. In: *Journal of Computational and Applied Mathematics* 125.1 (2000), pp. 171–182.

[19] J. Castro. “The Kolmogorov infinite dimensional equation in a Hilbert space via deep learning methods”. In: *Journal of Mathematical Analysis and Applications* 527.2 (2023), p. 127413.

[20] J. Chessari et al. “Numerical methods for backward stochastic differential equations: a survey”. In: *Probability Surveys* 20 (2023).

[21] PlanetMath contributors. *Compactly supported continuous functions are dense in L^p*. 2006.

[22] J. B. Conway. *A course in functional analysis*. New York: Springer, 2007.

[23] C. Cuchiero, P. Schmocker, and J. Teichmann. *Global universal approximation of functional input maps on weighted spaces*. arXiv preprint. 2023.

[24] W. E, J. Han, and A. Jentzen. “Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations”. In: *Communications in Mathematics and Statistics* 5 (2017), pp. 349–380.

[25] D. Filipović. *Consistency problems for Heath-Jarrow-Morton interest rate models*. Vol. 1760. Lecture Notes in Mathematics. Springer, 2001.

[26] D. Filipović, S. Tappe, and J. Teichmann. *Jump-diffusions in Hilbert spaces: existence, stability and numerics*. arXiv preprint. 2010.

[27] D. Filipović, S. Tappe, and J. Teichmann. “Jump-diffusions in Hilbert spaces: existence, stability and numerics”. In: *Stochastics: An International Journal of Probability and Stochastic Processes* 82.5 (2010), pp. 475–520.

[28] L. Gawarecki and V. Mandrekar. *Stochastic differential equations in infinite dimensions: with applications to stochastic partial differential equations*. Probability and Its Applications. Springer Berlin Heidelberg, 2011.

[29] D. Goodair and D. Crisan. *Stochastic calculus in infinite dimensions and SPDEs*. arXiv preprint arXiv:2203.17206. 2022.

[30] S. H. Haji and A. M. Abdulazeez. “Comparison of optimization techniques based on gradient descent algorithm: a review”. In: *PalArch’s Journal of Archaeology of Egypt/Egyptology* 18.4 (2021), pp. 2715–2743.

[31] P. Hall. "One-dimensional stable distributions (Translations of Mathematical Monographs 65)". In: *Bulletin of the London Mathematical Society* 19.5 (1987), pp. 504–506.

[32] J. Han, A. Jentzen, and W. E. "Solving high-dimensional partial differential equations using deep learning". In: *Proceedings of the National Academy of Sciences* 115 (2018), pp. 8505–8510.

[33] J. Han and J. Long. "Convergence of the deep BSDE method for coupled FBSDEs". In: *Probability, Uncertainty and Quantitative Risk* 5 (2020), p. 5.

[34] N. Hofmann and P. Mathé. "On quasi-Monte Carlo simulation of stochastic differential equations". In: *Mathematics of Computation* 66.218 (1997), pp. 573–589.

[35] K. Hornik, M. Stinchcombe, and H. White. "Multilayer feedforward networks are universal approximators". In: *Neural Networks* 2.5 (1989), pp. 359–366.

[36] K. Hornik, M. Stinchcombe, and H. White. "Multilayer feedforward networks are universal approximators". In: *Neural Networks* 2.5 (1989), pp. 359–366.

[37] C. Huré, H. Pham, and X. Warin. *Deep backward schemes for high-dimensional nonlinear PDEs.* arXiv preprint. 2020.

[38] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes.* Vol. 288. Grundlehren der Mathematischen Wissenschaften. Springer, 2003.

[39] I. T. Jolliffe and J. Cadima. "Principal component analysis: a review and recent developments". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2065 (2016), p. 20150202.

[40] L. Kapllani and L. Teng. *A backward differential deep learning-based algorithm for solving high-dimensional nonlinear backward stochastic differential equations.* arXiv preprint. 2024.

[41] D. P. Kingma and J. Ba. *Adam: a method for stochastic optimization.* arXiv preprint. 2017.

[42] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations.* Berlin, Heidelberg: Springer, 1992.

[43] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations.* Berlin: Springer, 1992.

[44] S. Koekebakker and F. Ollmar. "Forward curve dynamics in the Nordic electricity market". In: *Norwegian School of Economics and Business Administration* (2001).

[45] B. Negyesi, K. Andersson, and C. W. Oosterlee. "The one step Malliavin scheme: new discretization of BSDEs implemented with deep learning regressions". In: *IMA Journal of Numerical Analysis* 44 (2024), pp. 3595–3647.

[46] B. Negyesi, Z. Huang, and C. W. Oosterlee. *Generalized convergence of the deep BSDE method: a step towards fully-coupled FBSDEs and applications in stochastic control.* arXiv preprint. 2025.

[47] N. Nishikawa et al. “Two-layer neural network on infinite dimensional data: global optimization guarantee in the mean-field regime”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 32612–32623.

[48] T. D. H. van Nuland. “Noncompact uniform universal approximation”. In: *Neural Networks* 173 (2024), p. 106181.

[49] B. Øksendal. “Stochastic differential equations”. In: *Stochastic differential equations: an introduction with applications*. Springer, 2003, pp. 38–50.

[50] A. Pazy. *Semigroups of linear operators and applications to partial differential equations*. Vol. 44. Applied Mathematical Sciences. Springer, 1983.

[51] S. Peszat and J. Zabczyk. *Stochastic partial differential equations with Lévy noise: an evolution equation approach*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2007.

[52] K. Sato. *Lévy processes and infinitely divisible distributions*. Cambridge University Press, 1999.

[53] R. Schatten. *Norm ideals of completely continuous operators*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer, 1960.

[54] R. Schatten. *Norm ideals of completely continuous operators*. Springer, 1960.

[55] P. J. C. Spreij. “Stochastic integration”. In: *Lecture notes, University of Amsterdam* (2011).

[56] *Stochastic Models for Prices Dynamics in Energy and Commodity Markets: An Infinite-Dimensional Perspective*. Cham: Springer, 2023.

[57] A. V. Terekhov. “Generating the Laguerre expansion coefficients by solving a one-dimensional transport equation”. In: *Numerical Algorithms* 89.1 (2022), pp. 303–322.

[58] Y. Tian, Y. Zhang, and H. Zhang. “Recent advances in stochastic gradient descent in deep learning”. In: *Mathematics* 11.3 (2023), p. 682.

[59] H. Wang et al. *Deep learning-based BSDE solver for Libor market model with application to Bermudan swaption pricing and hedging*. arXiv preprint. 2018.