Quantization-based Bermudan option pricing in the FX world

Jean-Michel Fayolle * Vincent Lemaire † Thibaut Montes *†
Gilles Pagès †

March 3, 2021

Abstract

This paper proposes two numerical solution based on Product Optimal Quantization for the pricing of Bermudan options on Foreign Exchange (FX). More precisely, Bermudan Power Reverse Dual Currency options, where we take into account stochastic domestic and foreign interest rates on top of stochastic FX rate, hence we consider a 3-factor model. For these two numerical methods, we give an estimation of the L^2-error induced by such approximations and we illustrate them with market-based examples that highlight the speed of such methods.

Keywords— Foreign Exchange rates; Bermudan Options; Numerical method; Power Reverse Dual Currency; Product Optimal Quantization.

Introduction

Persistent low levels of interest rates in Japan in the latter decades of the 20th century were one of the core sources that led to the creation of structured financial products responding to the need of investors for coupons higher than the low yen-based ones. This started with relatively simple dual currency notes in the 80s where coupons were linked to foreign (i.e. non yen-based) currencies enabling payments of coupons significantly higher. As time (and issuers' competition) went by, such structured notes were iteratively "enhanced" to reverse dual currency, power reverse dual currency (PRDC), cancellable power reverse dual currency etc., each version adding further features such as limits, early repayment options, etc. Finally, in the early 2000s, the denomination xPRD took root to describe those structured notes typically long-dated (over 30y initial term) and based on multiple currencies (see [Wys17]). The total notional invested in such notes is likely to be in the hundreds of billions of USD. The valuation of such investments obviously requires the modeling of the main components driving the key risks, namely the interest rates of each pair of currencies involved as well as the corresponding exchange rates. In its simplest and most popular version, that means 3 sources of risk: domestic and foreign rates and the exchange rate. The 3-factor model discussed herein is an answer to that problem.

Gradually, as the note's features became more and more complex, further refinements to the modeling were needed, for instance requiring the inclusion of the volatility smile, the dependence

*The Independent Calculation Agent, ICA, 112 Avenue Kleber, 75116 Paris, France.
†Sorbonne Université, Laboratoire de Probabilités, Statistique et Modélisation, LPSM, Campus Pierre et Marie Curie, case 158, 4 place Jussieu, F-75252 Paris Cedex 5, France.

of implied volatilities on both the expiry and the strike[1] of the option, prevalent in the FX options market. Such more complete modeling should ideally consist in successive refinements of the initial modeling enabling consistency across the various flavors of xPRDs at stake.

The model discussed herein was one of the answers popular amongst practitioners for multiple reasons: it was accounting for the main risks – interest rates in the currencies involved and exchange rates – in a relatively simple manner and the numerical implementations proposed at that time were based on simple extensions of well-known single dimensional techniques such as 3 dimensional trinomial trees, PDE based method (see [Pit05]) or on Monte Carlo simulations.

Despite the qualities of these methods, the calculation time could be rather slow (around 20 minutes with a trinomial tree for one price), especially when factoring in the cost for hedging (that is, measuring the sensitivities to all the input parameters) and even more post 2008, where the computation of risk measures and their sensitivities to market values became a central challenge for the financial markets participants. Indeed, even though these products were issued towards the end of the 20th century, they are still present in the banks's books and need to be considered when evaluating counterparty risk computations such as Credit Valuation Adjustment (CVA), Debt Valuation Adjustment (DVA), Funding Valuation Adjustment (FVA), Capital Valuation Adjustment (KVA), ..., in short xVA's (see [BMP13, CBB14, Gre15] for more details on the subject). Hence, a fast and accurate numerical method is important for being able to produce the correct values in a timely manner. The present paper aims at providing an elegant and efficient answer to that problem of numerical efficiency based on Optimal Quantization. Our novel method allows us reach a computation time of 1 or 2 seconds at the expense of a systematic error that we quantify in Section 3.

Let $P(t,T)$ be the value at time t of one unit of the currency delivered (that is, paid) at time T, also known as a zero coupon price or discount factor. A few iterations were needed by researchers and practitioners before the seminal family of Heath-Jarrow-Morton models came about. The general Heath-Jarrow-Morton (HJM) family of yield curve models can be expressed as follows – although originally expressed by its authors in terms of rates dynamics, the two are equivalent, see [HJM92] – in a n-factor setting, we have for the curve $P(t,T)$ that

$$\frac{dP(t,T)}{P(t,T)} = r_t dt + \sum_i \sigma_i\big(t,T,P(t,T)\big) dW_t^i$$

where r_t is the (stochastic) instantaneous rate at time t barrer (therefore a random variable), W^i, $i = 1, \cdots, n$ are n correlated Brownian motions and $\sigma_i\big(t,T,P(t,T)\big)$ are volatility functions in the most general settings (with the obvious constraint that $\sigma_i\big(T,T,P(T,T)\big) = 0$). Indeed, the general HJM framework allows for the volatility functions $\sigma_i\big(t,T,P(t,T)\big)$ to also depend on the yield curve's (random) levels up to t – actually through forward rates – and therefore be random too. However, it has been demonstrated in [EKMV92] that, to keep a tractable version (i.e. a finite number of state variables), the volatility functions must be of a specific form, namely, of the mean-reverting type (where the mean reversion can also depend on time). We use this way of expressing the model as a mean to recall that such model is essentially the usual and well-known Black Scholes model applied to all and any zero-coupon prices, with various enhancements regarding number of factors and volatility functions, to keep the calculations tractable. For further details and theory, one can refer to some of the following articles [EKFG96, EKMV92, HJM92, BS73]. Of course, such a framework can be applied to any yield curve. In its simplest form (i.e. flat volatility and one-factor), we have under the risk-neutral measure

$$\frac{dP(t,T)}{P(t,T)} = r_t dt + \sigma(T-t) dW_t \tag{0.1}$$

[1]In the case of the FX, the implied volatility is expressed in function of the delta.

where W is a standard Brownian motion under the risk-neutral probability. In that case, σ is the flat volatility, which means the volatility of (zero-coupon) interest rates. That is often referred to as a Hull-White model without mean reversion (see [HW93]) or a continuous-time version of the Ho-Lee model. In the rest of the paper, we work with the model presented in (0.1) for the diffusion of the zero coupon although the extension to non-flat volatilities is easily feasible.

About the Foreign Exchange (FX) rate, we denote by S_t the value at time $t > 0$ of one unit of foreign currency in the domestic currency. Its dynamics is driven by a Black-Scholes diffusion model with the following equation

$$\frac{dS_t}{S_t} = (r_t^d - r_t^f)dt + \sigma_S dW_t^S$$

where r_t^d and r^f are the instantaneous rates at time t of the domestic and foreign currency respectively, both supposed to be stochastic, σ_S is the (deterministic) volatility of the FX rate and W^S is a standard Brownian motion under the domestic risk-neutral probability.

Let us briefly recall the principle of the adopted numerical method, Optimal quantization. Optimal Quantization is a numerical method whose aim is to approximate optimally, for a given norm, a continuous random signal by a discrete one with a given cardinality at most N. [She97] was the first to work on it for the uniform distribution on unit hypercubes. Since then, it has been extended to more general distributions with applications to Signal transmission in the 50's at the Bell Laboratory (see [GG82]). Formally, let Z be an $\mathbb{R}^d$-valued random vector with distribution $\mathbb{P}_Z$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that $Z \in L^2(\mathbb{P})$. We search for Γ_N, a finite subset of $\mathbb{R}^d$ defined by $\Gamma_N := \{z_1^N, \ldots, z_N^N\} \subset \mathbb{R}^d$, solution to the following problem

$$\min_{\Gamma_N \subset \mathbb{R}^d, |\Gamma_N| \leqslant N} \|Z - \widehat{Z}^N\|_2$$

where $\widehat{Z}^N$ denotes the nearest neighbour projection of Z onto Γ_N. This problem can be extended to the L^p-optimal quantization by replacing the L^2-norm by the L^p-norm but this not in the scope of this paper. In our case, we mostly consider quadratic one-dimensional optimal quantization, i.e $d = 1$ and $p = 2$. The existence of an optimal quantizer at level N goes back to [CAGM97] (see also [Pag98, GL00] for further developments). In the one-dimensional case, if the distribution of Z is absolutely continuous with a *log-concave* density, then there exists a unique optimal quantizer at level N, see [Kie83]. We scale to the higher dimension using Optimal Product Quantization which deals with multi-dimensional quantizers built by considering the cartesian product of one-dimensional optimal quantizers.

Considering again $Z = (Z^\ell)_{\ell=1:d}$, a $\mathbb{R}^d$-valued random vector. First, we look separately at each component Z^ℓ independently by building a one-dimensional optimal quantization $\widehat{Z}^\ell$ of size N^ℓ, with quantizer $\Gamma_\ell^{N_\ell} = \big\{z_{i_\ell}^\ell, i_\ell \in \{1, \cdots, N_\ell\}\big\}$ and then, by applying the cartesian product between the one-dimensional optimal quantizers, we build the product quantizer $\Gamma^N = \prod_{\ell=1}^d \Gamma_\ell^{N_\ell}$ with cardinality $N = N^1 \times \cdots \times N^d$ by

$$\Gamma^N = \big\{(z_{i_1}^1, \cdots, z_{i_\ell}^\ell, \cdots, z_{i_d}^d), \quad i_\ell \in \{1, \cdots, N_\ell\}, \quad \ell \in \{1, \cdots, d\}\big\}.$$

Then, in the 1990s, [Pag98] developed quantization-based cubature formulas for numerical integration purposes and expectation approximations. Indeed, let f be a continuous function $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ such that $f(Z) \in L^1(\mathbb{P})$, we can define the following quantization-based cubature formula using the discrete property of the quantizer $\widehat{Z}^N$

$$\mathbb{E}\left[f(\widehat{Z}^N)\right] = \sum_{i=1}^N p_i f(z_i^N)$$

where $p_i = \mathbb{P}(\widehat{Z}^N = z_i^N)$. Then, one could want to approximate $\mathbb{E}\big[f(Z)\big]$ by $\mathbb{E}\big[f(\widehat{Z}^N)\big]$ when the first expression cannot be computed easily. For example, this case is exactly the problem one encounters when trying to price European options. We know the rate of convergence of the weak error induced by this cubature formula, i.e $\exists \alpha \in (0,2]$, depending on the regularity of f such that

$$\lim_{N\to+\infty} N^{\alpha} \big| \mathbb{E}\big[f(Z)\big] - \mathbb{E}\big[f(\widehat{Z}^N)\big]\big| \leqslant C_{f,X} < +\infty.$$

For more results on the rate of convergence, the value of α, we refer to [Pag18] for a survey in $\mathbb{R}^d$ and to [LMP19] for recent improved results in the one-dimensional case.

Later on, in a series of papers, among them [BP03, BPP05] extended this method to the computation of conditional expectations allowing to deals with nonlinear problems in finance and, more precisely, to the pricing and hedging of American/Bermudan options, which is the part we are interested in. These problems are of the form

$$\sup_{\tau} \mathbb{E}\left[\mathrm{e}^{-\int_0^{\tau} r_s^d ds}\, \psi_{\tau}(S_{\tau}) \right]$$

where $\big(\mathrm{e}^{-\int_0^{t_k} r_s^d ds}\, \psi_{t_k}(S_{t_k}) \big)_{k=0,\dots,n}$ is the obstacle function and $\tau : \Omega \to \{t_0, t_1, \dots, t_n\}$ is a stopping time for the filtration $(\mathcal{F}_{t_k})_{k\geqslant 0}$ where $\mathcal{F}_t = \sigma\big(S_s, P^d(s,T), P^f(s,T), s \leqslant t\big)$ is the natural filtration to consider because the foreign exchange rate and the zero-coupon curves are observables in the market.

In this paper, we will present two numerical solutions, motivated by the works described above, to the problem of the evaluation of Bermudan option on Foreign Exchange rate with stochastic interest rates. The paper is organised as follows. First, in Section 1, we introduce the diffusion models for the zero coupon curves and the foreign exchange rate we work with. In Section 2, we describe in details the financial product we want to evaluate: Bermudan option on foreign exchange rate. In this Section, we express the *Backward Dynamic Programming Principle* and study the regularity of the obstacle process and the value function. Then, in Section 3, we propose two numerical solutions for pricing the financial product defined above based on Product Quantization and we study the L^2-error induced by these numerical approximations. In Section 4, several examples are presented in order to compare the two methods presented in Section 3. First, we begin with plain European option, this test is carried out in order to benchmark the methods because a closed-form formula is known for the price of a European Call/Put in the 3-factor model. Then, we compare the two methods in the case of a Bermudan option with several exercise dates. Finally, in Appendix A, we make some change of numéraire and in Appendix B, we give the closed-form formula for the price of an European Call, in the 3-factor model, used in Section 4 as a benchmark.

1 Diffusion Models

Interest Rate Model. We shall denote by $P(t,T)$ the value at time t of one unit of the currency delivered (that is, paid) at time T, also known as a zero coupon price or discount factor. When t is today, this function can usually be derived from the market price of standard products, such as bonds and interest rate swaps in the market, along with an interpolation scheme (for the dates different than the maturities of the market rates used). In a simple single-curve framework, the derivation of the initial curve, that is, the zero coupons $P(0,T)$ for $T > 0$ is rather simple, through relatively standard methods of curve stripping. In more enhanced frameworks accounting for multiple yield curves such as having different for curves for discounting and forward rates, those methods are somewhat more demanding but still relatively straightforward. We focus herein on the simple single-curve framework.

In our case we are working with financial products on Foreign Exchange (FX) rates between the domestic and the foreign currency, hence we will be working with zero coupons in the domestic currency denoted by $P^d(t,T)$ and zero coupons in the foreign currency denoted by $P^f(t,T)$. The diffusion of the domestic zero-coupon curve under the domestic risk-neutral probability $\mathbb{P}$ is given by

$$\frac{dP^d(t,T)}{P^d(t,T)} = r_t^d dt + \sigma_d(T-t)dW_t^d$$

where W^d is a $\mathbb{P}$-Brownian Motion, r_t^d is the domestic instantaneous rate at time t and σ_d is the volatility for the domestic zero coupon curve. For the foreign zero-coupon curve, the diffusion is given, under the foreign neutral probability $\widetilde{\mathbb{P}}$, by

$$\frac{dP^f(t,T)}{P^f(t,T)} = r_t^f dt + \sigma_f(T-t)d\widetilde{W}_t^f$$

where $\widetilde{W}^f$ is a $\widetilde{\mathbb{P}}$-Brownian Motion, r_t^f is the foreign instantaneous rate at time t and σ_f is the volatility for the foreign zero coupon curve. Both volatilities σ_d and σ_f are positive real constants. The two probabilities $\widetilde{\mathbb{P}}$ and $\mathbb{P}$ are supposed to be equivalent, i.e $\widetilde{\mathbb{P}} \sim \mathbb{P}$ and it exists ρ_{df} defined as limit of the quadratic variation $\langle W^d, \widetilde{W}^f \rangle_t = \rho_{df} t$.

Remark. Such a framework to model random yield curves has been quite popular with practitioners due to its elegance, simplicity and intuitive understanding of rates dynamics through time yet providing a comprehensive and consistent modelling of an entire yield curve through time. Indeed, it is mathematically and numerically easily tractable. It carries no path dependency and allows the handling of multiple curves for a given currency as well as multiple currencies – and their exchange rates – as well as equities (when one wishes to account for random interest rates). It allows negative rates and can be refined by adding factors (Brownian motions).

However, it cannot easily cope with smile or non-normally distributed shocks or with internal curve "oddities" or specifics such as different volatilities for different swap tenors within the same curve dynamics. Nonetheless, our aim being to propose a model and a numerical method which make possible to produce risk computations (such as xVA's) in an efficient way, these properties are of little importance. That said, when it comes to deal with accounting for random rates in long-dated derivatives valuations, its benefits far outweigh its limitations and its use for such applications is popular, see [NP14] for the pricing of swaptions, [Pit05] for PRDCs...

Foreign Exchange Model. The diffusion of the foreign exchange (FX) rate defined under the domestic risk-neutral probability is

$$\frac{dS_t}{S_t} = (r_t^d - r_t^f)dt + \sigma_S dW_t^S$$

where W_t^S is a $\mathbb{P}$-Brownian Motion under the domestic risk-neutral probability such that their exist ρ_{Sd} and $\rho_{Sf} \in [-1,1]$ defined as limit of the quadratic variations $\langle W^S, W^d \rangle_t = \rho_{Sd} t$ and $\langle W^S, \widetilde{W}^f \rangle_t = \rho_{Sf} t$, respectively.

Finally, the processes, expressed in the domestic risk-neutral probability $\mathbb{P}$, are

$$\begin{cases} \dfrac{dP^d(t,T)}{P^d(t,T)} = r_t^d dt + \sigma_d(T-t)dW_t^d \\ \dfrac{dS_t}{S_t} = (r_t^d - r_t^f)dt + \sigma_S dW_t^S \\ \dfrac{dP^f(t,T)}{P^f(t,T)} = \big(r_t^f - \rho_{Sf}\sigma_S\sigma_f(T-t)\big)dt + \sigma_f(T-t)dW_t^f \end{cases} \tag{1.1}$$

where W^f, defined by $dW_s^f = d\widetilde{W}_s^f + \rho_{Sf}\sigma_S ds$, is a $\mathbb{P}$-Brownian motion, as shown in Appendix A. Note at this stage, that the zero-coupons $P^d(0,t)$ and $P^f(0,t)$ are the quoted prices at time 0 in their respective markets of one unit of domestic and foreign currency. In particular these are deterministic quantities since they are are observable at time 0.

Using Itô's formula, we can explicitly express the processes

$$\begin{cases} P^d(t,T) = P^d(0,T)\exp\left(\displaystyle\int_0^t \left(r_s^d - \frac{\sigma_d^2(T-s)^2}{2}\right)ds + \sigma_d\int_0^t (T-s)dW_s^d\right) \\ S_t = S_0\exp\left(\displaystyle\int_0^t \left(r_s^d - r_s^f - \frac{\sigma_S^2}{2}\right)ds + \sigma_S W_t^S\right) \\ P^f(t,T) = P^f(0,T)\exp\left(\displaystyle\int_0^t \left(r_s^f - \rho_{Sf}\sigma_S\sigma_f(T-s) - \frac{\sigma_f^2(T-s)^2}{2}\right)ds + \sigma_f\int_0^t (T-s)dW_s^f\right) \end{cases}.$$

From these equations, we deduce $\exp\Big(-\int_0^t r_s^d ds\Big)$ and $\exp\Big(-\int_0^t r_s^f ds\Big)$, by taking $T=t$ and using that $P^d(t,t) = P^f(t,t) = 1$, it follows that

$$\begin{cases} \exp\left(-\displaystyle\int_0^t r_s^d ds\right) = \varphi_d(t)\exp\left(\sigma_d\displaystyle\int_0^t (t-s)dW_s^d\right) \\ \exp\left(-\displaystyle\int_0^t r_s^f ds\right) = \varphi_f(t)\exp\left(\sigma_f\displaystyle\int_0^t (t-s)dW_s^f\right), \end{cases}$$

where

$$\varphi_d(t) = P^d(0,t)\exp\left(-\sigma_d^2\int_0^t \frac{(t-s)^2}{2}ds\right)$$

and

$$\varphi_f(t) = P^f(0,t)\exp\left(-\int_0^t \left(\rho_{Sf}\sigma_S\sigma_f(t-s) + \frac{\sigma_f^2(t-s)^2}{2}\right)ds\right).$$

These expressions for the domestic and the foreign discount factors will be useful in the following sections of the paper.

2 Bermudan options

2.1 Product Description

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be our domestic risk neutral probability space. We want to evaluate the price of a Bermudan option with maturity $T > 0$ on the FX rate $(S_t)_{t\geqslant 0}$ with payoffs $\psi_{t_k}(S_{t_k})$ which can be exercised by its owner at predetermined dates $t_0 = 0 < t_1 < \cdots < t_k < \cdots t_n = T$. The FX rate $(S_t)_{t\geqslant 0}$ is defined by

$$S_t = \frac{1}{\exp\Big(-\int_0^t r_s^d ds\Big)} S_0 \varphi_f(t) \exp\Big(-\frac{\sigma_S^2}{2}t + \sigma_S W_t^S + \sigma_f \int_0^t (t-s) dW_s^f\Big)$$

with

$$\exp\Big(-\int_0^t r_s^d ds\Big) = \varphi_d(t) \exp\Big(\sigma_d \int_0^t (t-s) dW_s^d\Big).$$

The payoff functions ψ_{t_k} are non-negative Borel functions satisfying

$$\forall\, k = 0, \ldots, n, \quad \mathbb{E}\left[\psi_{t_k}(S_{t_k})^2\right] < +\infty.$$

At a given time t, the *observable assets* in the market are the foreign exchange rate S_t and the zero-coupon curves $\big(P^d(t,T)\big)_{T\geqslant t}$ and $\big(P^f(t,T)\big)_{T\geqslant t}$, hence the natural filtration to be considered is

$$\mathcal{F}_t = \sigma\big(\mathcal{N}_{\mathbb{P}}, S_s, P^d(s,T), P^f(s,T), s \leqslant t\big) = \sigma\big(\mathcal{N}_{\mathbb{P}}, W_s^S, W_s^d, W_s^f, s \leqslant t\big)$$

where $\mathcal{N}_{\mathbb{P}}$ denotes the $\mathbb{P}$–negligible sets in $\mathcal{A}$. Let $\tau : \Omega \to \{t_0, t_1, \ldots, t_n\}$ be a stopping time for $(\mathcal{F}_{t_k})_{k\geqslant 0}$ and $\mathcal{T} = \mathcal{T}^{\mathcal{F}}$ the set of all $[0,T]$-valued $(\mathcal{F}_{t_k})_{k\geqslant 0}$-stopping times. In this paper, we consider problems where the horizon is finite then we define $\mathcal{T}_k^n$, the set of $(\mathcal{F}_{t_k})$-stopping times having a.s. values in $\{t_k, \ldots, t_n\}$.

Hence, the price at time t_k of the Bermudan option is given by $(\mathcal{F}_{t_k}, \mathbb{P})$-Snell envelope of the *obstacle* process $\big(e^{-\int_0^{t_k} r_s^d ds} \psi_{t_k}(S_{t_k})\big)$ at time t_k, namely

$$V_k = \sup_{\tau \in \mathcal{T}_k^n} \mathbb{E}\left[\, \mathrm{e}^{-\int_0^\tau r_s^d ds}\, \psi_\tau(S_\tau) \mid \mathcal{F}_{t_k}\,\right]$$

and V_k is called the *Snell envelope* of the obstacle process $\big(\,\mathrm{e}^{-\int_0^{t_k} r_s^d ds}\, \psi_{t_k}(S_{t_k})\big)_{0\leqslant k\leqslant n}$.

Remark. The financial products we consider in the applications are PRDC. Their payoffs (see Figure 1) have the following expression

$$\psi_{t_k}(x) = \min\left(\max\left(\frac{C_f(t_k)}{S_0}x - C_d(t_k), \mathrm{Floor}(t_k)\right), \mathrm{Cap}(t_k)\right) \tag{2.1}$$

where $\mathrm{Floor}(t_k)$, $\mathrm{Cap}(t_k)$, $C_f(t_k)$ and $C_d(t_k)$ are known constants at time 0. More precisely, $\mathrm{Floor}(t_k)$ and $\mathrm{Cap}(t_k)$ are the floor and cap values, and $C_f(t_k)$ and $C_d(t_k)$ are the coupons value we wish to compare to the foreign and the domestic currency, respectively.

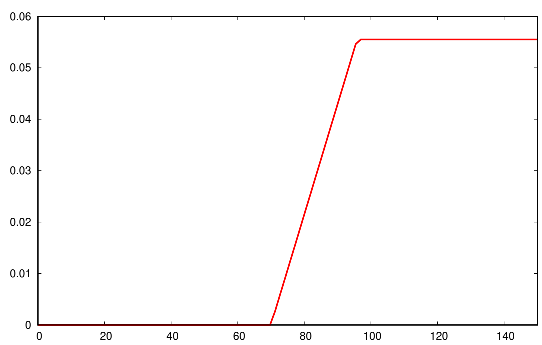


Figure 1: *Example of a PRDC payoff* $\psi_{t_k}(S_{t_k}) = \min\left(\left(0.189\frac{S_{t_k}}{88.17} - 0.15\right)_+, 0.0555\right)$ *at time* t_k.

The interesting feature of such functions is that their (right) derivative have a compact support.

2.2 Backward Dynamic Programming Principle

As a discrete time $(\mathcal{F}_{t_k}, \mathbb{P})$-Snell envelope, the sequence $(V_k)_{0\leqslant k\leqslant n}$ satisfies the so-called *Backward Dynamic Programming Principle* (*BDPP*) i.e. the recursive formula:

$$\begin{cases} V_n = \mathrm{e}^{-\int_0^{t_n} r_s^d ds}\,\psi_{t_n}(S_{t_n}), \\ V_k = \max\Big(\mathrm{e}^{-\int_0^{t_k} r_s^d ds}\,\psi_{t_k}(S_{t_k}), \mathbb{E}[V_{k+1} \mid \mathcal{F}_{t_k}]\Big), \qquad 0 \leqslant k \leqslant n-1 \end{cases}. \tag{2.2}$$

To make this backward recursion numerically tractable we have to solve several problems. The first (and main) one is to "Markovianize" the above (*BDPP*) by writing the payoffs $\psi_{t_k}(S_{t_k})$ as functions of an $(\mathcal{F}_{t_k})_{k=0,\dots,n}$-Markov chain. To this end, we first notice that the obstacle process $\big(\mathrm{e}^{-\int_0^t r_s^d ds}\,\psi_t(S_t)\big)_{t=t_0,\dots,t_n}$ can be rewritten (in fact for every $t\in[0,T]$) as a function h_t of two processes X_t and Y_t such that

$$h_t(X_t, Y_t) = \mathrm{e}^{-\int_0^t r_s^d ds}\,\psi_t(S_t)$$

where

$$h_t(x,y) = \varphi_d(t)\,\mathrm{e}^{-y}\,\psi_t\bigg(S_0\frac{\varphi_f(t)}{\varphi_d(t)}\,\mathrm{e}^{-\sigma_S^2 t/2+x+y}\bigg) \tag{2.3}$$

and (X,Y) is defined by

$$(X_t, Y_t) = \left(\sigma_S W_t^S + \sigma_f\int_0^t (t-s)dW_s^f, -\sigma_d\int_0^t (t-s)dW_s^d\right)$$

so that it is clearly $\mathcal{F}_t$-adapted. However it is clear that the couple $(X_{t_k}, Y_{t_k})_{k=0,\dots,n}$ is not an $(\mathcal{F}_{t_k})$-Markov process and we need to add extra components to achieve our Markovianization.

Several choices are *a priori* possible to perform this re-parametrization the payoff process. We turned to the obviously $(\mathcal{F}_{t_k})_{k=0,\dots,n}$-adapted 4-tuple $(X, Y, W^f, W^d)_{t=t_0,\dots,t_n}$, which is in fact an $(\mathcal{F}_{t_k})_k$-Markov process as can be easily checked (see (2.4) later).

This choice is motivated by the following – slightly paradoxal – argument: the (non-Markov) 2-dimensional model process (X,Y) turns out to be very close to the Markovian model (X, Y, W^f, W^d) if one has in mind that these two Brownian motions are "weighted" in our formulas by σ^f and σ^d which are very small in practice (few basis points ($1\,\mathrm{bps} = 10^{-4}$)). This will in turn make reasonable our non-Markovian approximation consisting, as a second step (see Section 3.3, in forcing the Markov property in the backward dynamic programming principle by directly conditioning w.r.t. (X_{t_k}, Y_{t_k}) to speed up the numerical procedure). A more financial justification could be that these quantities also appear naturally in the discount factor and in S_t.

From now on, in order to alleviate notations, we denote by $X_k = X_{t_k}$, $W_k^f = W_{t_k}^f$, $Y_k = Y_{t_k}$, $W_k^d = W_{t_k}^d$, $W_k^S = W_{t_k}^S$ and $h_k = h_{t_k}$.

The $(\mathcal{F}_{k_k})_{k=0,\dots,n}$-Markov property of $(X_{t_k}, W_{t_k}^f, Y_{t_k}, W_{t_k}^d)_{k=0,\dots,n}$ is a consequence of the following decomposition

$$
\begin{cases}
X_{k+1} = X_k + \sigma_f \mathbf{h} W_k^f + \sigma_S \displaystyle\int_{t_k}^{t_{k+1}} dW_s^S + \sigma_f \int_{t_k}^{t_{k+1}} (t_{k+1} - s) dW_s^f, \\
W_{k+1}^f = W_k^f + \displaystyle\int_{t_k}^{t_{k+1}} dW_s^f, \\
Y_{k+1} = Y_k - \sigma_d \mathbf{h} W_k^d - \sigma_d \displaystyle\int_{t_k}^{t_{k+1}} (t_{k+1} - s) dW_s^d, \\
W_{k+1}^d = W_k^d + \displaystyle\int_{t_k}^{t_{k+1}} dW_s^d.
\end{cases} \tag{2.4}
$$

using that (W^S, W^f, W^s) is a process with stationary independent increments. Setting $\mathbf{h} = \frac{T}{n}$ this can be written as

$$
\begin{cases}
X_{k+1} = X_k + \sigma_f \mathbf{h} W_k^f + G_{k+1}^1 \\
W_{k+1}^f = W_k^f + G_{k+1}^2 \\
Y_{k+1} = Y_k - \sigma_d \mathbf{h} W_k^d + G_{k+1}^3 \\
W_{k+1}^d = W_k^d + G_{k+1}^4,
\end{cases} \tag{2.5}
$$

where the increments $(G_k)_{k=1,\ldots,n}$ are i.i.d., centered, normally distributed like G defined by

$$
G = \begin{pmatrix} G^1 \\ G^2 \\ G^3 \\ G^4 \end{pmatrix} \sim \mathcal{N}\Big(0, \Sigma_{\mathbf{h}}\Big) \tag{2.6}
$$

with $\Sigma_{\mathbf{h}} = \Big[\mathbb{C}\mathrm{ov}\left(G^i, G^j\right) \Big]_{i,j=1:4}$ given by the (symmetric) matrix

$$
\Sigma_{\mathbf{h}} = \mathbf{h} \begin{bmatrix}
\sigma_S^2 + \mathbf{h}(\frac{\sigma_f^2}{3}\mathbf{h} + \frac{\sigma_S \sigma_f \rho_{Sf}}{2}) & \sigma_S \rho_{Sf} + \frac{\sigma_f}{2}\mathbf{h} & -\mathbf{h}\big(\frac{\sigma_S \sigma_d \rho_{Sd}}{2} + \frac{\sigma_f \sigma_d \rho_{df}}{3}\mathbf{h}\big) & \sigma_S \rho_{Sd} + \frac{\sigma_f \rho_{df}}{2}\mathbf{h} \\
\times & 1 & -\frac{\sigma_d \rho_{df}}{2}\mathbf{h} & \rho_{df} \\
\times & \times & \frac{\sigma_d^2}{3}\mathbf{h}^2 & -\frac{\sigma_d}{2}\mathbf{h} \\
\times & \times & \times & 1
\end{bmatrix}. \tag{2.7}
$$

It follows from Equation (2.5) that the Markov chain $(X_{t_k}, W_{t_k}^f, Y_{t_k}, W_{t_k}^d)_{k=0,\ldots,n}$ is homogeneous with transition reading on Borel test-functions $f : \mathbb{R}^4 \to \mathbb{R}$,

$$
Pf(x, u, y, v) = \mathbb{E}\big[f(x + \sigma_f \mathbf{h} u + G^1, u + G^2, y - \sigma_d \mathbf{h} v + G^3, v + G^4)\big]. \tag{2.8}
$$

Then, it is classical background that using the $(\mathcal{F}_{t_k})_k$-Markov property of (X_k, W_k^f, Y_k, W_k^d), the *BDPP* (2.2) can be written as follows,

$$
\begin{cases}
V_n = h_n(X_n, Y_n), \\
V_k = \max\Big(h_k(X_k, Y_k), \mathbb{E}\left[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\right]\Big), \qquad 0 \leqslant k \leqslant n-1.
\end{cases} \tag{2.9}
$$

and $V_k = v_k(X_k, W_k^f, Y_k, W_k^d)$ for some Borel functions $v_k : \mathbb{R}^4 \to \mathbb{R}$ satisfying $\forall x, w^f, y, w^d \in \mathbb{R}$,

$$
\begin{cases}
v_n(x, w^f, y, w^d) = h_n(x, y), \\
v_k(x, w^f, y, w^d) = \max\Big(h_k(x, y), P v_{k+1}(x, w^f, y, w^d)\Big), \quad 0 \leqslant k \leqslant n-1.
\end{cases} \tag{2.10}
$$

Payoff regularity. First, we look at the regularity of the payoff. The next proposition will then allow us to study the regularity of the value function through the propagation of the local Lipschitz property by the transition of the Markov process.

Proposition 2.1. *If ψ_{t_k}, $k \in \{0, \dots, n\}$, are Lipschitz continuous with Lipschitz coefficients $[\psi_{t_k}]_{Lip}$ with compactly supported (right) derivatives (such as the payoff defined in (2.1)) then $h_k(x, y)$ given by (2.3) is locally Lipschitz continuous, there exist real constants $C_k^\psi > A_k^\psi > 0$ only depending on ψ_{t_k} and on its right derivative ψ'_{t_k} such that*

$$\forall x, x', y, y' \in \mathbb{R}, \quad |h_k(x, y) - h_k(x', y')| \leqslant \mathrm{e}^{|y| \vee |y'|} \big(A_k^\psi |x - x'| + C_k^\psi |y - y'|\big).$$

Proof. Let g_k be defined by

$$g_k(x, y) = \psi_{t_k}\left(S_0 \frac{\varphi_f(t_k)}{\varphi_d(t_k)} \mathrm{e}^{-\sigma_S^2 t_k/2 + x + y}\right).$$

As ψ'_{t_k} has a compact support, then it exists $c \in \mathbb{R}$ such that

$$|(\psi_{t_k}(\mathrm{e}^x))'| = |\,\mathrm{e}^x \psi'_{t_k}(\mathrm{e}^x)| \leqslant \|\psi'_{t_k}\|_\infty \sup_{x \in \operatorname{supp} \psi'_{t_k}} \mathrm{e}^x \leqslant \|\psi'_{t_k}\|_\infty \,\mathrm{e}^c .$$

Hence

$$|g_k(x, y) - g_k(x', y')| \leqslant \frac{A_k^\psi}{\varphi_d(t_k)} \big(|x - x'| + |y - y'|\big)$$

with $A_k^\psi = [\psi_{t_k}]_{Lip} S_0 \varphi_f(t_k) \,\mathrm{e}^{-\sigma_S^2 t_k/2} \,\|\psi'_{t_k}\|_\infty \,\mathrm{e}^c$. Then for every $x, x', y, y' \in \mathbb{R}$, we have

$$\begin{aligned}
|h_k(x, y) - h_k(x', y')| &= \varphi_d(t_k) \big|\,\mathrm{e}^{-y} g_k(x, y) - \mathrm{e}^{-y'} g_k(x', y')\big| \\
&\leqslant \varphi_d(t_k) \Big(\big|\,\mathrm{e}^{-y} g_k(x, y) - \mathrm{e}^{-y'} g_k(x, y)\big| + \big|\,\mathrm{e}^{-y'} g_k(x, y) - \mathrm{e}^{-y'} g_k(x', y')\big|\Big) \\
&\leqslant \varphi_d(t_k) \Big(\big|\,\mathrm{e}^{-y} - \mathrm{e}^{-y'}\big| \cdot \|\psi_{t_k}\|_\infty + \mathrm{e}^{-y'} \big|g_k(x, y) - g_k(x', y')\big|\Big) \\
&\leqslant \mathrm{e}^{|y| \vee |y'|} \Big(\varphi_d(t_k) \|\psi_{t_k}\|_\infty |y - y'| + A_k^\psi \big(|x - x'| + |y - y'|\big)\Big).
\end{aligned}$$

The result follows with $C_k^\psi = A_k^\psi + \varphi_d(t_k) \|\psi_{t_k}\|_\infty$. ☐

The next Lemma shows that the transition of the Markov process propagates and controls the local Lipschitz continuity of a function f. This result will be helpful to estimate the error induced by the numerical approximation (2.10).

Lemma 2.2. *Let P denote the Markov transition operator defined by $Pf(x, u, y, v) = \mathbb{E}\big[f(x + \sigma_f \mathbf{h} u + G^1, u + G^2, y - \sigma_d \mathbf{h} v + G^3, v + G^4)\big]$ with $G = (G^\ell)_{1 \leqslant \ell \leqslant 4} \sim \mathcal{N}(0, \Sigma_{\mathbf{h}})$, $\Sigma_{\mathbf{h}}$ given by(2.7). If the function f satisfies the following local Lipschitz property,*

$$|f(x, u, y, v) - f(x', u', y', v')| \leqslant \big(A|x - x'| + B|u - u'| + C|y - y'| + D|v - v'|\big) \times \mathrm{e}^{|y| \vee |y'| + b|v| \vee |v'|}$$

then

$$\begin{aligned}
|Pf(x, u, y, v) - Pf(x', u', y', v')| \leqslant \big(&A|x - x'| + (B + A\sigma_f \mathbf{h})|u - u'| \\
&+ C|y - y'| + (D + C\sigma_d \mathbf{h})|v - v'|\big) \\
&\times \bar{\kappa}(b) \,\mathrm{e}^{|y| \vee |y'| + (b + \sigma_d \mathbf{h})|v| \vee |v'|}
\end{aligned} \tag{2.11}$$

with $\bar{\kappa}(b) = \mathbb{E}\big[\exp(|G^3| + b|G^4|)\big] < +\infty$.

Proof. From Jensen's inequality and our assumption on f

$$
\begin{aligned}
&|Pf(x,u,y,v) - Pf(x',u',y',v')| \\
&\quad \leqslant \mathbb{E}\Big[\big|f(x+\sigma_f \mathbf{h} u + G^1, u+G^2, y-\sigma_d \mathbf{h} v + G^3, v+G^4) \\
&\qquad\qquad - f(x'+\sigma_f \mathbf{h} u' + G^1, u'+G^2, y'-\sigma_d \mathbf{h} v' + G^3, v'+G^4)\big|\Big] \\
&\quad \leqslant \big(A|x-x'| + (B+A\sigma_f \mathbf{h})|u-u'| + C|y-y'| + (D+C\sigma_d \mathbf{h})|v-v'|\big) \\
&\qquad\qquad \times \mathrm{e}^{|y|\vee|y'| + (b+\sigma_d \mathbf{h})|v|\vee|v'|}\, \mathbb{E}\big[\mathrm{e}^{|G^3|+b|G^4|}\big].
\end{aligned}
$$

□

Value function regularity. If the functions $(\psi_{t_k})_{0\leqslant k\leqslant n}$ are defined by Equation (2.1), then the value function preserves a local Lipschitz property at each time t_k. More precisely we have the following Lemma.

Lemma 2.3. *If the functions ψ_{t_k}, $k \in \{0,\dots,n\}$, are Lipschitz continuous with Lipschitz coefficient $[\psi_{t_k}]_{Lip}$ with a compactly supported (right) derivative, then for each $k \in \{0,\dots,n\}$, v_k defined by* (2.10) *is locally Lipschitz continuous, and there exist $\bar{A}_k$, $\bar{B}_k$, $\bar{C}_k$, $\bar{D}_k$ positive constants (with closed forms given by* (2.14) *and* (2.15) *below) and $b_k = \sigma_d \mathbf{h}(n-k)$, $k = 0,\dots,n$ such that, for every $x,x',u,u',y,y',v,v' \in \mathbb{R}$, both*

$$
\begin{aligned}
&|v_k(x,u,y,v) - v_k(x',u',y',v')| \qquad (2.12)\\
&\qquad \leqslant \big(\bar{A}_k|x-x'| + \bar{B}_k|u-u'| + \bar{C}_k|y-y'| + \bar{D}_k|v-v'|\big)\, \mathrm{e}^{|y|\vee|y'| + b_k|v|\vee|v'|},
\end{aligned}
$$

and

$$
\begin{aligned}
&|Pv_{k+1}(x,u,y,v) - Pv_{k+1}(x',u',y',v')| \qquad (2.13)\\
&\qquad \leqslant \big(\bar{A}_k|x-x'| + \bar{B}_k|u-u'| + \bar{C}_k|y-y'| + \bar{D}_k|v-v'|\big)\, \mathrm{e}^{|y|\vee|y'| + b_k|v|\vee|v'|}.
\end{aligned}
$$

Proof. For every $x,x',u,u',y,y',v,v' \in \mathbb{R}$,

$$
\begin{aligned}
|v_n(x,u,y,v) - v_n(x',u',y',v')| &\leqslant \big(\bar{A}_n|x-x'| + \bar{B}_n|u-u'| + \bar{C}_n|y-y'| + \bar{D}_n|v-v'|\big) \\
&\qquad \times \mathrm{e}^{|y|\vee|y'| + b_n|v|\vee|v'|}
\end{aligned}
$$

where

$$
\bar{A}_n = A_n^{\psi}, \quad \bar{B}_n = 0, \quad \bar{C}_n = C_n^{\psi}, \quad \bar{D}_n = 0, \quad \bar{b}_n = 0
$$

where A_n^{ψ} and C_n^{ψ} come from Proposition 2.1. Using now Lemma 2.2 recursively and the elementary inequality $\max(a, b+c) \leqslant \max(a,b) + c$ (as $x \mapsto \max(a,x)$ is 1-Lipschitz), we have

$$
\begin{aligned}
&|v_k(x,u,y,v) - v_k(x',u',y',v')| \\
&\qquad \leqslant \max\big(|h_k(x,y) - h_k(x',y')|, |Pv_{k+1}(x,u,y,v) - Pv_{k+1}(x',u',y',v')|\big),
\end{aligned}
$$

so that

$$
\begin{aligned}
&|v_k(x,u,y,v) - v_k(x',u',y',v')| \\
&\quad \leqslant \max\Bigg(\mathrm{e}^{|y|\vee|y'|}\big(A_k^{\psi}|x-x'| + C_k^{\psi}|y-y'|\big), \\
&\qquad\qquad \big(\bar{A}_{k+1}|x-x'| + (\bar{B}_{k+1} + \bar{A}_{k+1}\sigma_f \mathbf{h})|u-u'| + \bar{C}_{k+1}|y-y'| \\
&\qquad\qquad + (\bar{D}_{k+1} + \bar{C}_{k+1}\sigma_d \mathbf{h})|v-v'|\big)\bar{\kappa}(b_{k+1})\, \mathrm{e}^{|y|\vee|y'| + (b_{k+1}+\sigma_d \mathbf{h})|v|\vee|v'|}\Bigg).
\end{aligned}
$$

One derives by a backward induction that $b_k = \sigma_d \mathbf{h}(n-k)$, and denoting $\bar{\pi}_k = \prod_{j=0}^{k} \bar{\kappa}(b_j)$ (with $\bar{\kappa}(b) = \mathbb{E}\big[\exp(|G^3| + b|G^4|)\big]$)

$$\bar{A}_k = \max_{k\leqslant \ell \leqslant n} \left(A_\ell^\psi \frac{\bar{\pi}_\ell}{\bar{\pi}_k} \right), \qquad \bar{B}_k = \sigma_f\, \mathbf{h} \frac{1}{\bar{\pi}_k} \sum_{\ell=k+1}^{n} \bar{\pi}_\ell \bar{A}_\ell, \tag{2.14}$$

and

$$\bar{C}_k = \max_{k\leqslant \ell \leqslant n} \left(C_\ell^\psi \frac{\bar{\pi}_\ell}{\bar{\pi}_k} \right), \qquad \bar{D}_k = \sigma_f\, \mathbf{h} \frac{1}{\bar{\pi}_k} \sum_{\ell=k+1}^{n} \bar{\pi}_\ell \bar{C}_\ell. \tag{2.15}$$

Finally, one checks that (2.13) holds as well in view of the respective values appearing inside the max, thanks to Lemma 2.2. □

3 Bermudan pricing using Optimal Quantization

In this section, we propose two numerical solutions based on Product Optimal Quantization for the pricing of Bermudan options on the FX rate S_t. First, we remind briefly what an optimal quantizer is and what we mean by a product quantization tree. Second, we present a first numerical solution, based on quantization of the Markovian 4-tuple (X, W^f, Y, W^d), to solve the numerical problem (2.9) and detail the L^2-error induced by this approximation. However, remember that we are looking for a method that makes possible to compute xVA's risk measures in a reasonable time but this solution can be too time consuming in practice due to the dimensionality of the quantized processes. That is why we present a second numerical solution which reduces the dimensionality of the problem by considering an approximate problem, based on quantization of the non-Markovian couple (X, Y), introducing a systematic error induced by the non-markovianity and we study the L^2-error produced by this approximation.

3.1 Theoretical background (one-dimensional case).

The aim of Optimal Quantization is to produce optimal spatial discretization of random vector (or its distribution) at a given *level* N i.e. to determine a finite Γ_N, a set with cardinality at most N, which minimises the L^p-approximation error among all such sets Γ. We consider for our needs only the one-dimensional case. Let $Z \in L^p_{\mathbb{R}}$, $p \in [1, +\infty)$ be a random variable with distribution $\mathbb{P}_Z$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Let $\Gamma = \{z_1, \dots, z_N\} \subset \mathbb{R}$ be a subset of size N, called N-quantizer. We may assume up to renumbering that $z_1 < z_2 < \cdots < z_N$. Composing Z with any Borel nearest neighbour projection $\mathrm{Proj}_\Gamma : \mathbb{R} \to \{z_1, \dots, z_N\}$, i.e.

$$\widehat{Z}^\Gamma = \mathrm{Proj}_\Gamma(Z)$$

achives the best pointwise approximation of Z by a Γ-valued random variable. In particular

$$\mathrm{Proj}_\Gamma = \sum_{i=1}^{N} z_i \, \mathbb{1}_{Z \in C_i(\Gamma)}$$

where $\{C_i(\Gamma) : 1 \leqslant i \leqslant N\}$ is a partition induced of $\mathbb{R}^d$ satisfying

$$C_i(\Gamma) \subset \Big\{ \xi \in \mathbb{R}, |\xi - z_i| \leqslant \min_{j \neq i} |\xi - z_j| \Big\}.$$

is a nearest neighbour projection. Then the *L^p-mean quantization error* induced by the quantizer $\widehat{Z}^\Gamma$ only depends on Γ and $\mathbb{P}_Z$ is defined by

$$\|\mathrm{dist}(Z,\Gamma)\|_p = \|Z - \widehat{Z}^\Gamma\|_p = \left(\mathbb{E}\left[\min_{i\in\{1,\cdots,N\}} |Z - z_i|^p\right]\right)^{1/p}. \tag{3.1}$$

It is convenient to define the L^p-distortion function at level N as the pth power of the L^p-mean quantization error on $(\mathbb{R})^N$:

$$\mathcal{Q}_{p,N}(z_1,\ldots,z_N) := \mathbb{E}\left[\min_{i\in\{1,\cdots,N\}} |Z - z_i|^p\right] = \|Z - \widehat{Z}^{\{z_1,\ldots,z_N\}}\|_p^p.$$

It has been shown that this function attains a minimum at an N-tuple $z^{(N)} = (z_1,\ldots,z_N)$ producing an L^p-optimal quantizer $\Gamma_N = \{z_i, i = 1,\cdots,N\}$ (see e.g. [Pag18] and the references therein for details). For our purpose, we need to compute such optimal grids for the normal distribution when $p = 2$: this has been done and made available on the website

`www.quantize.math-fi.com`

A really interesting and useful property concerning quadratic optimal quantizers is the stationarity property.

Proposition 3.1. *(Stationarity) Assume that the support of $\mathbb{P}_X$ has at least N elements. Any L^2-optimal N-quantizer $\Gamma_N \in (\mathbb{R})^N$ is stationary in the following sense: for every Voronoï quantization $\widehat{X}^N$ of X,*

$$\mathbb{E}\big[X \mid \widehat{X}^N\big] = \widehat{X}^N.$$

The rate of decay of the minimal L^p-mean quantization error is a natural question and useful for numerical results (references can be found in [Pag18]) but in our case we need more: the L^s-convergence rate of the L^s-quantization error induced by a sequence of L^p-optimal quantizers $p < s$. This problem, known as the distortion mismatch, was first addressed in [GLP08] and then generalized in [PS18, Theorem 4.3]. The following theorem is a simplified version of the original one in [PS18].

Theorem 3.2. *[L^r-L^s-distortion mismatch] Let $Z : (\Omega,\mathcal{A},\mathbb{P}) \to \mathbb{R}$ be a random variable and let $r \in (0,+\infty)$. Assume that the distribution $\mathbb{P}_Z$ of Z has a non-zero absolutely continuous component with respect to the Lebesgue measure. Let $r \geqslant 1$ and let $(\Gamma_N)_{N\geqslant 1}$ be a sequence of L^r-optimal N-quantizers. Assume that $Z \in \bigcap_{p\geqslant 1} L^p(\mathbb{P})$. Then*

$$\forall\, s \in (0, r+1), \qquad \limsup_N N\|Z - \widehat{Z}^N\|_s < +\infty.$$

Product Quantization. Now, let $Z = (Z^\ell)_{\ell=1:d}$ be an $\mathbb{R}^d$-valued random vector with distribution $\mathbb{P}_Z$ defined on a probability space $(\Omega,\mathcal{A},\mathbb{P})$. There are two approaches if one wishes to scale to higher dimensions. Either one applies the above framework directly to the random vector Z and build an optimal quantizer of Z, or one may consider separately each component Z^ℓ independently, build a one-dimensional optimal quantization $\widehat{Z}^\ell$, of size N^ℓ, with quantizer $\Gamma_\ell^{N^\ell} = \big\{z_{i_\ell}^\ell, i_\ell \in \{1,\cdots,N^\ell\}\big\}$ and then build the product quantizer $\Gamma^N = \prod_{\ell=1}^d \Gamma_\ell^{N^\ell}$ of size $N = N^1 \times \cdots \times N^d$ defined by

$$\Gamma^N = \big\{(z_{i_1}^1,\cdots,z_{i_\ell}^\ell,\cdots,z_{i_d}^d),\quad i_\ell \in \{1,\cdots,N_\ell\},\quad \ell \in \{1,\cdots,d\}\big\}.$$

In our case we chose the second approach. Indeed, it is much more flexible when dealing with normal distribution, like in our case. We do not need to solve the d-dimensional minimization problem at each time step. We only need to load precomputed optimal quantizer of standard normal distribution $\mathcal{N}(0,1)$ and then take advantage of the stability of optimal quantization by rescaling in one dimension in the sense that if $\Gamma^N = \{z_i, 1 \leqslant i \leqslant N\}$ is optimal at level N for $\mathcal{N}(0,1)$ then $\mu + \sigma\Gamma^N$ (with obvious notations) is optimal for $\mathcal{N}(\mu, \sigma^2)$.

Even though there exist fast methods for building optimal quantizers in two-dimension based on deterministic methods like in the one-dimensional case, when dealing with optimal quantization of bivariate Gaussian vector, we may face numerical instability when the covariance matrix is ill-conditioned: so is the case if the variance of one coordinate is relatively high compared to the second one (which is our case in this paper). This a major drawback as we are looking for a fast numerical solution able to produce prices in a few seconds and this is possible when using product optimal quantization.

Quantization Tree. Now, instead of considering a random variable Z, let $(Z_t)_{t\in[0,T]}$ be a stochastic process following a Stochastic Differential Equation (SDE)

$$Z_t = Z_0 + \int_0^t b_s(Z_s)ds + \int_0^t \sigma(s, Z_s)dW_s$$

with $Z_0 = z_0 \in \mathbb{R}^d$, W a standard Brownian motion living on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and b and σ satisfy the standard assumptions in order to ensure the existence of a strong solution of the SDE.

What we call Quantization Tree is defined, for chosen time steps $t_k = Tk/n, k = 0, \cdots, n$, by quantizers $\widehat{Z}_k$ of Z_k (Product Quantizers in our case) at dates t_k and the transition probabilities between date t_k and date t_{k+1}. Although $(\widehat{Z}_k)_k$ is no longer a Markov process we will consider the transition probabilities $\pi_{ij}^k = \mathbb{P}(\widehat{Z}_{k+1} = z_j^{k+1} \mid \widehat{Z}_k = z_i^k)$. We can apply this methodology because, with the model we consider, we know all the marginal laws of our processes at each date of interest.

In the next subsection, we present the approach based on the quantization tree previously defined that allows us to approximate the price of Bermudan options where the risk factors are driven by the 3-factor model (1.1).

3.2 Quantization tree approximation: Markov case

Our first approach in order to spatially discretize (2.9) is to quantize (X_k, W_k^f, Y_k, W_k^d) at each instant t_k by *product quantization* i.e. using quadratic optimal quantizations $\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d$ of their marginals X_k, W_k^f, Y_k and W_k^d of size N_k^X, $N_k^{W^f}$, N_k^Y and $N_k^{W^d}$ respectively. Then $(\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)_{k=0,\ldots,n}$ is no longer a Markov process. We "force" in some sense the (lost) Markov property in the Backward Dynamic Programming Principle by introducing a *Quantized Backward Dynamic Programming Principle* (QBDPP) as follows

$$\begin{cases} \widehat{V}_n^{(\mathrm{M})} = h_n(\widehat{X}_n, \widehat{Y}_n), \\ \widehat{V}_k^{(\mathrm{M})} = \max\Big(h_k(\widehat{X}_k, \widehat{Y}_k), \mathbb{E}\big[\widehat{V}_{k+1}^{(\mathrm{M})} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\big]\Big), \qquad 0 \leqslant k \leqslant n-1. \end{cases} \tag{3.2}$$

Moreover, let $(x_{i_1}^k)_{i_1=1:N_k^X}$, $(u_{i_2}^k)_{i_2=1:N_k^{W^f}}$, $(y_{i_3}^k)_{i_3=1:N_k^Y}$ and $(v_{i_4}^k)_{i_4=1:N_k^{W^d}}$ be the associated centroids of $\widehat{X}_k$, $\widehat{W}_k^f$, $\widehat{Y}_k$ and $\widehat{W}_k^d$ respectively. Using the discrete property of the optimal quantizers, the conditional expectation appearing in (3.2) can be rewritten as

$$\begin{aligned} &\mathbb{E}\big[\widehat{V}_{k+1}^{(\mathrm{M})} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) = \big(x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k\big)\big] \\ &\qquad = \mathbb{E}\big[\widehat{v}_{k+1}^{(\mathrm{M})}(\widehat{X}_{k+1}, \widehat{W}_{k+1}^f, \widehat{Y}_{k+1}, \widehat{W}_{k+1}^d) \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) = \big(x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k\big)\big] \\ &\qquad = \sum_{j_1,j_2,j_3,j_4} \pi_{i,j}^{(\mathrm{M}),k}\, \widehat{v}_{k+1}^{(\mathrm{M})}\big(x_{j_1}^{k+1}, u_{j_2}^{k+1}, y_{j_3}^{k+1}, v_{j_4}^{k+1}\big) \end{aligned}$$

where $\pi_{i,j}^{(\mathrm{M}),k}$, with $i = (i_1, i_2, i_3, i_4)$ and $j = (j_1, j_2, j_3, j_4)$, is the conditional probability defined by

$$\pi_{i,j}^{(\mathrm{M}),k} = \mathbb{P}\Big(\big(\widehat{X}_{k+1}, \widehat{W}^f_{k+1}, \widehat{Y}_{k+1}, \widehat{W}^d_{k+1}\big) = \big(x^{k+1}_{j_1}, u^{k+1}_{j_2}, y^{k+1}_{j_3}, v^{k+1}_{j_4}\big) \\ \mid \big(\widehat{X}_k, \widehat{W}^f_k, \widehat{Y}_k, \widehat{W}^d_k\big) = \big(x^k_{i_1}, u^k_{i_2}, y^k_{i_3}, v^k_{i_4}\big)\Big). \tag{3.3}$$

This corresponds to marginal quantization tree as introduced in [BPP05] (and named in [PPP04]). We are interested in the error induced by the above algorithm and more precisely its L^2-error. Standard error bounds from [BPP05] need to be adapted since the payoff functions are not Lipschitz.

Theorem 3.3. *Let the Markov transition $Pf(x, u, y, v)$ defined in (2.8) be locally Lipschitz in the sense of Lemma 2.2. Assume that all the payoff functions $(\psi_{t_k})_{0\leqslant k\leqslant n}$ are Lipschitz continuous with compactly supported (right) derivative. Then the L^2-error between $\widehat{V}^{(\mathrm{M})}$ defined by (3.2) and V defined by (2.9) using the quantization approximation $(\widehat{X}_k, \widehat{W}^f_k, \widehat{Y}_k, \widehat{W}^d_k)$ is upper-bounded by*

$$\big\|V_k - \widehat{V}^{(\mathrm{M})}_k\big\|_2^2 \leqslant \sum_{\ell=k}^{n} C_{X_\ell}\big\|X_\ell - \widehat{X}_\ell\big\|_{2p}^2 + C_{Y_\ell}\big\|Y_\ell - \widehat{Y}_\ell\big\|_{2p}^2 + C_{W^d_\ell}\big\|W^d_\ell - \widehat{W}^d_\ell\big\|_{2p}^2 + C_{W^f_\ell}\big\|W^f_\ell - \widehat{W}^f_\ell\big\|_{2p}^2, \tag{3.4}$$

where $1 < p < 3/2$ and $q \geqslant 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$ and

$$\begin{aligned} C_{X_\ell} &= 2K^q_\ell(0)(A^\psi_\ell)^2 + 4K^q_\ell(b_\ell)\bar{A}^2_\ell, & C_{W^d_\ell} &= 4\bar{B}^2_\ell K^q_\ell(b_\ell), \\ C_{Y_\ell} &= 2K^q_\ell(0)(C^\psi_\ell)^2 + 4K^q_\ell(b_\ell)\bar{C}^2_\ell, & C_{W^f_\ell} &= 4\bar{D}^2_\ell K^q_\ell(b_\ell) \end{aligned} \tag{3.5}$$

with the function K^q_ℓ defined by

$$\forall b > 0, \quad K^q_\ell(b) = \big\|\, \mathrm{e}^{|Y_\ell|\vee|\widehat{Y}_\ell| + b|W^d_\ell|\vee|\widehat{W}^d_\ell|}\,\big\|_{2q}^2. \tag{3.6}$$

As a consequence if $\bar{N} = \min_k N_k$, we have

$$\lim_{\bar{N}\to+\infty} \big\|V_k - \widehat{V}^{(\mathrm{M})}_k\big\|_2^2 = 0. \tag{3.7}$$

Remark. From the definition of the processes X_k, W^f_k, Y_k and W^d_k, all are Gaussian random variables, hence the functions K^q_k, $q \geqslant 1$, $k \in \{0, \dots, n\}$, introduced in (3.6) are well defined. Indeed, let $Z \sim \mathcal{N}(0, \sigma_Z)$ be a Gaussian random variable with variance σ_Z^2 and $\widehat{Z}$ an optimal quadratic quantizer of Z with cardinality N then $\forall\, \lambda \in \mathbb{R}_+$

$$\forall q \geqslant 1, \big\|\, \mathrm{e}^{\lambda|Z|\vee|\widehat{Z}|}\,\big\|_{2q} = \Big(\mathbb{E}\big[\, \mathrm{e}^{2q\lambda|Z|\vee|\widehat{Z}|}\,\big]\Big)^{\frac{1}{2q}} \leqslant \Big(2\,\mathbb{E}\big[\, \mathrm{e}^{2q\lambda|Z|}\,\big]\Big)^{\frac{1}{2q}} \leqslant 2^{\frac{1}{2q}}\, \mathrm{e}^{q^2\lambda^2\sigma_Z^2}$$

since, by the stationarity property of $\widehat{Z}$ and the convexity of $e^{2q\lambda\cdot}$, $\mathbb{E}\big[\, \mathrm{e}^{2q\lambda|\widehat{Z}|}\,\big] \leqslant \mathbb{E}\big[\, \mathrm{e}^{2q\lambda|Z|}\,\big]$.

Proof. The error between the Snell envelope and its approximation is given by

$$|V_k - \widehat{V}^{(\mathrm{M})}_k| \leqslant \max\Big(\big|h_k(X_k, Y_k) - h_k(\widehat{X}_k, \widehat{Y}_k)\big|, \\ \big|\,\mathbb{E}\big[V_{k+1} \mid (X_k, W^f_k, Y_k, W^d_k)\big] - \mathbb{E}\big[\widehat{V}^{(\mathrm{M})}_{k+1} \mid (\widehat{X}_k, \widehat{W}^f_k, \widehat{Y}_k, \widehat{W}^d_k)\big]\big|\Big)$$

thus, using the local Lipschitz property of h_k established in Proposition 2.1 and Hölder's inequality with $p, q \geqslant 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$, the L^2-error is upper-bounded by

$$
\begin{aligned}
\big\|V_k - \widehat{V}_k^{(\mathrm{M})}\big\|_2^2 &\leqslant \big\|h_k(X_k, Y_k) - h_k(\widehat{X}_k, \widehat{Y}_k)\big\|_2^2 \\
&\qquad + \big\| \mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big] - \mathbb{E}\big[\widehat{V}_{k+1}^{(\mathrm{M})} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\big]\big\|_2^2 \\
&\leqslant 2\big\| \mathrm{e}^{|Y_k| \vee |\widehat{Y}_k|} \big\|_{2q}^2 \Big((C_k^\psi)^2 \big\|Y_k - \widehat{Y}_k\big\|_{2p}^2 + (A_k^\psi)^2 \big\|X_k - \widehat{X}_k\big\|_{2p}^2 \Big) \\
&\qquad + \big\| \mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big] - \mathbb{E}\big[\widehat{V}_{k+1}^{(\mathrm{M})} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\big]\big\|_2^2.
\end{aligned}
\tag{3.8}
$$

Looking at the last term, we have

$$
\begin{aligned}
&\mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big] - \mathbb{E}\big[\widehat{V}_{k+1}^{(\mathrm{M})} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\big] \\
&\qquad = \mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big] - \mathbb{E}\big[V_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\big] \\
&\qquad\quad + \mathbb{E}\big[V_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\big] - \mathbb{E}\big[\widehat{V}_{k+1}^{(\mathrm{M})} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\big].
\end{aligned}
$$

Now, we inspect the L^2-error of each term on the right-hand side of the equality.

- For the first term, we notice that

$$
\begin{aligned}
&\mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big] - \mathbb{E}\big[V_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\big] \\
&\qquad\qquad = Pv_{k+1}(X_k, W_k^f, Y_k, W_k^d) - Pv_{k+1}(\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)
\end{aligned}
$$

then by Lemma 2.3 we obtain

$$
\begin{aligned}
&\big| \mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big] - \mathbb{E}\big[V_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\big]\big| \\
&\leqslant \big(\bar{A}_k|X_k - \widehat{X}_k| + \bar{B}_k|W_k^f - \widehat{W}_k^f| + \bar{C}_k|Y_k - \widehat{Y}_k| + \bar{D}_k|W_k^d - \widehat{W}_k^d|\big)\, \mathrm{e}^{|Y_k| \vee |\widehat{Y}_k| + b_k|W_k^d| \vee |\widehat{W}_k^d|}.
\end{aligned}
$$

Hence, using Hölder's inequality with $p, q \geqslant 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$,

$$
\begin{aligned}
&\big\| \mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big] - \mathbb{E}\big[V_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\big]\big\|_2^2 \\
&\leqslant 4\Big(\bar{A}_k^2\big\|X_k - \widehat{X}_k\big\|_{2p}^2 + \bar{B}_k^2\big\|W_k^f - \widehat{W}_k^f\big\|_{2p}^2 + \bar{C}_k^2\big\|Y_k - \widehat{Y}_k\big\|_{2p}^2 + \bar{D}_k^2\big\|W_k^d - \widehat{W}_k^d\big\|_{2p}^2\Big) \\
&\qquad \times \big\| \mathrm{e}^{|Y_k| \vee |\widehat{Y}_k| + b_k|W_k^d| \vee |\widehat{W}_k^d|} \big\|_{2q}^2.
\end{aligned}
\tag{3.9}
$$

- The last one is useful for the induction, indeed

$$
\big\| \mathbb{E}\big[V_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\big] - \mathbb{E}\big[\widehat{V}_{k+1}^{(\mathrm{M})} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\big]\big\|_2^2 \leqslant \big\|V_{k+1} - \widehat{V}_{k+1}^{(\mathrm{M})}\big\|_2^2. \tag{3.10}
$$

Combining (3.8), (3.9) and (3.10) and using the function K_k^q defined in (3.6) we get

$$
\begin{aligned}
\big\|V_k - \widehat{V}_k^{(\mathrm{M})}\big\|_2^2 &\leqslant 2K_k^q(0)\Big((C_k^\psi)^2\big\|Y_k - \widehat{Y}_k\big\|_{2p}^2 + (A_k^\psi)^2\big\|X_k - \widehat{X}_k\big\|_{2p}^2\Big) \\
&\qquad + 4K_k^q(b_k)\Big(\bar{A}_k^2\big\|X_k - \widehat{X}_k\big\|_{2p}^2 + \bar{B}_k^2\big\|W_k^f - \widehat{W}_k^f\big\|_{2p}^2 \\
&\qquad\qquad + \bar{C}_k^2\big\|Y_k - \widehat{Y}_k\big\|_{2p}^2 + \bar{D}_k^2\big\|W_k^d - \widehat{W}_k^d\big\|_{2p}^2\Big) + \big\|V_{k+1} - \widehat{V}_{k+1}^{(\mathrm{M})}\big\|_2^2.
\end{aligned}
$$

By induction we deduce (3.4) with constants defined in (3.5).

Finally, using the L^r-L^s mismatch Theorem 3.2 applied with $r = 2$ and $s = 2p > r$ for the four L^{2p}-quantization error terms related to X, Y, W^f and W^d yields, if $1 < p < 3/2$, we obtain (3.7). □

To conclude this section, although considering product optimal quantizer in four dimensions for (X_k, W_k^f, Y_k, W_k^d) seems to be natural, the computational cost associated to the resulting QBDPP is too high, of order $O(n \times (\max N_k)^2)$. Moreover the computation of the transition probabilities needed for the evaluation of the terms $\mathbb{E}\big[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\big]$ is challenging. These transition probabilities cannot be computed using deterministic numerical integration methods and we have to use Monte Carlo estimators. Even though it is feasible, it is a drawback for the method since it increases drastically the computation time for calibrating the quantization tree. In the next section we provide a solution to these problems which consists in reducing the dimension of the problem at the price of adding a systematic error, which turns out to be quite small in practice.

3.3 Quantization tree approximation: Non Markov case

In this part, we want to reduce the dimension of the problem in order to scale down the numerical complexity of the pricer. For that we discard the processes W^d and W^f in the quantization tree and only keep X and Y. Doing so, we lose the Markovian property of our original model but we drastically reduce the numerical complexity of the problem. Thence, (2.9) is still approximated by forcing the Markov property in the Dynamic programming this with respect to $(\widehat{X}_k, \widehat{Y}_k)$:

$$\begin{cases} \widehat{V}_n^{(\text{NM})} = h_n(\widehat{X}_n, \widehat{Y}_n), \\ \widehat{V}_k^{(\text{NM})} = \max\Big(h_k(\widehat{X}_k, \widehat{Y}_k), \mathbb{E}\big[\widehat{V}_{k+1}^{(\text{NM})} \mid (\widehat{X}_k, \widehat{Y}_k)\big]\Big), \qquad 0 \leqslant k \leqslant n-1 \end{cases} \tag{3.11}$$

where for every $k = 0, \dots, n$, $\widehat{X}_k$ and $\widehat{Y}_k$ are quadratic optimal quantizations of X_k and Y_k of size N_k^X and N_k^Y respectively. Let $N_k = N_k^X \times N_k^Y$ denotes the size of the product grid. Let $(x_{i_1}^k)_{i_1=1:N_k^X}$ and $(y_{i_3}^k)_{i_3=1:N_k^Y}$ be the associated centroids of $\widehat{X}_k$ and $\widehat{Y}_k$ respectively. Again, as in the Markovian case, using the discrete property of the optimal quantizers, the conditional expectation appearing in (3.11) can be rewritten as

$$\begin{aligned} \mathbb{E}\big[\widehat{V}_{k+1}^{(\text{NM})} \mid (\widehat{X}_k, \widehat{Y}_k) = \big(x_{i_1}^k, y_{i_2}^k\big)\big] &= \mathbb{E}\big[\widehat{v}_{k+1}^{(\text{NM})}(\widehat{X}_{k+1}, \widehat{Y}_{k+1}) \mid (\widehat{X}_k, \widehat{Y}_k) = \big(x_{i_1}^k, y_{i_2}^k\big)\big] \\ &= \sum_{j_1, j_2} \pi_{i,j}^{(\text{NM}),k} \, \widehat{v}_{k+1}\big(x_{j_1}^{k+1}, y_{j_2}^{k+1}\big) \end{aligned}$$

where $\pi_{i,j}^{(\text{NM}),k}$, with $i = (i_1, i_2)$ and $j = (j_1, j_2)$, is the conditional probability defined by

$$\pi_{i,j}^{(\text{NM}),k} = \mathbb{P}\Big(\big(\widehat{X}_{k+1}, \widehat{Y}_{k+1}\big) = \big(x_{j_1}^{k+1}, y_{j_2}^{k+1}\big) \mid \big(\widehat{X}_k, \widehat{Y}_k\big) = \big(x_{i_1}^k, y_{i_2}^k\big)\Big). \tag{3.12}$$

Theorem 3.4. *Let the Markov transition $Pf(x, u, y, v)$ be defined by (2.8) be locally Lipschitz in the sense of Lemma 2.2. Assume that all the payoff functions $(\psi_{t_k})_{0 \leqslant k \leqslant n}$ are Lipschitz continuous with compactly supported (right) derivative. Then the L^2-error between $\widehat{V}^{(\text{NM})}$ defined by (3.11) and V defined by (2.9) using the quantization approximation $(\widehat{X}_k, \widehat{Y}_k)$ is upper-bounded by*

$$\begin{aligned} \big\|V_k - \widehat{V}_k^{(\text{NM})}\big\|_2^2 \leqslant \sum_{\ell=k}^{n-1} \Big(& C_{W_\ell^f} \big\|W_\ell^f - \mathbb{E}[W_\ell^f \mid (X_\ell, Y_\ell)]\big\|_{2p}^2 + C_{W_\ell^d} \big\|W_\ell^d - \mathbb{E}[W_\ell^d \mid (X_\ell, Y_\ell)]\big\|_{2p}^2 \\ & + C_{X_\ell} \big\|X_\ell - \widehat{X}_\ell\big\|_{2p}^2 + C_{Y_\ell} \big\|Y_\ell - \widehat{Y}_\ell\big\|_{2p}^2 \Big) \end{aligned} \tag{3.13}$$

where $1 < p < 3/2$ and $q \geqslant 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$, moreover

$$\begin{aligned} C_{X_\ell} &= 2K_k^q(0)\big((A_k^\psi)^2 + \bar{A}_k^2\big), & C_{W_\ell^f} &= 2\widetilde{K}_k^q(b_k)\bar{B}_k^2, \\ C_{Y_\ell} &= 2K_k^q(0)\big((C_k^\psi)^2 + \bar{C}_k^2\big), & C_{W_\ell^d} &= 2\widetilde{K}_k^q(b_k)\bar{D}_k^2. \end{aligned}$$

with $K_k^q(0) = \|e^{|Y_k|\vee|\widehat{Y}_k|}\|_{2q}^2$ *and* $\widetilde{K}_k^q(b) = \|e^{|Y_k|+b|W_k^d|\vee|\,\mathbb{E}[W_k^d|(X_k,Y_k)]|}\|_{2q}^2$*. Taking the limit in* $\bar{N} = \min_k N_k$*, the size of the quadratic optimal quantizers, we have*

$$\limsup_{\bar{N}\to+\infty} \|V_k - \widehat{V}_k^{(\text{NM})}\|_2^2 \leqslant \mathcal{E}_n^{(\text{NM})}(k) \tag{3.14}$$

with $\mathcal{E}_n^{(\text{NM})}(k) = \sum_{\ell=k}^{n-1} C_{W_\ell^f}\|W_\ell^f - \mathbb{E}[W_\ell^f \mid (X_\ell, Y_\ell)]\|_{2p}^2 + C_{W_\ell^d}\|W_\ell^d - \mathbb{E}[W_\ell^d \mid (X_\ell, Y_\ell)]\|_{2p}^2$.

Proof. We apply the same methodology as in the proof for the Markov case. The error between the Snell envelope and its approximation is given by

$$|V_k-\widehat{V}_k^{(\text{NM})}| \leqslant \max\Big(\big|h_k(X_k,Y_k)-h_k(\widehat{X}_k,\widehat{Y}_k)\big|, \big|\,\mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big]-\mathbb{E}\big[\widehat{V}_{k+1}^{(\text{NM})} \mid (\widehat{X}_k,\widehat{Y}_k)\big]\big|\Big)$$

thus, using Proposition 2.1 and Hölder's inequality with $p, q \geqslant 1$ such that $\frac{1}{p}+\frac{1}{q} = 1$, the L^2-error is given by

$$\begin{aligned}
\|V_k - \widehat{V}_k\|_2^2 &\leqslant \|h_k(X_k,Y_k) - h_k(\widehat{X}_k,\widehat{Y}_k)\|_2^2 \\
&\qquad + \|\,\mathbb{E}\big[V_{k+1}^{(\text{NM})} \mid (X_k, W_k^f, Y_k, W_k^d)\big] - \mathbb{E}\big[\widehat{V}_{k+1} \mid (\widehat{X}_k,\widehat{Y}_k)\big]\|_2^2 \\
&\leqslant 2\|\,\mathrm{e}^{|Y_k|\vee|\widehat{Y}_k|}\,\|_{2q}^2\Big((C_k^\psi)^2\|Y_k - \widehat{Y}_k\|_{2p}^2 + (A_k^\psi)^2\|X_k - \widehat{X}_k\|_{2p}^2\Big) \\
&\qquad + \|\,\mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big] - \mathbb{E}\big[\widehat{V}_{k+1}^{(\text{NM})} \mid (\widehat{X}_k,\widehat{Y}_k)\big]\|_2^2.
\end{aligned} \tag{3.15}$$

The last term in Equation (3.15) can be decomposed as follows

$$\begin{aligned}
\mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big] &- \mathbb{E}\big[\widehat{V}_{k+1}^{(\text{NM})} \mid (\widehat{X}_k,\widehat{Y}_k)\big] \\
&= \mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big] - \mathbb{E}\big[V_{k+1} \mid (\widehat{X}_k,\widehat{Y}_k)\big] \\
&\quad + \mathbb{E}\big[V_{k+1} \mid (\widehat{X}_k,\widehat{Y}_k)\big] - \mathbb{E}\big[\widehat{V}_{k+1}^{(\text{NM})} \mid (\widehat{X}_k,\widehat{Y}_k)\big].
\end{aligned} \tag{3.16}$$

As $\widehat{X}_k, \widehat{Y}_k$ is a function of (X_k, Y_k) and $\sigma(X_k,Y_k) \subset \sigma(X_k, W_k^f, Y_k, W_k^d)$ we observe that

$$\mathbb{E}\big[\widehat{V}_{k+1}^{(\text{NM})} \mid (\widehat{X}_k,\widehat{Y}_k)\big] = \mathbb{E}\big[\,\mathbb{E}\big[\widehat{V}_{k+1}^{(\text{NM})} \mid (X_k, W_k^f, Y_k, W_k^d)\big] \mid (\widehat{X}_k,\widehat{Y}_k)\big].$$

Note that the two terms in the decomposition (3.16) are orthogonal in $L^2(\mathbb{P})$ by the definition of conditional expectation as an orthogonal projection. Consequently

$$\begin{aligned}
\|\,\mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big] &- \mathbb{E}\big[\widehat{V}_{k+1}^{(\text{NM})} \mid (\widehat{X}_k,\widehat{Y}_k)\big]\|_2^2 \\
&= \|\,\mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big] - \mathbb{E}\big[V_{k+1} \mid (\widehat{X}_k,\widehat{Y}_k)\big]\|_2^2 \\
&\quad + \|\,\mathbb{E}\big[V_{k+1} \mid (\widehat{X}_k,\widehat{Y}_k)\big] - \mathbb{E}\big[\widehat{V}_{k+1}^{(\text{NM})} \mid (\widehat{X}_k,\widehat{Y}_k)\big]\|_2^2 \\
&\leqslant \|\,\mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big] - \mathbb{E}\big[V_{k+1} \mid (\widehat{X}_k,\widehat{Y}_k)\big]\|_2^2 \\
&\quad + \|V_{k+1} - \widehat{V}_{k+1}^{(\text{NM})}\|_2^2.
\end{aligned}$$

Now it remains to control the term representative of the Markovian default, namely

$$\mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big] - \mathbb{E}\big[V_{k+1} \mid (\widehat{X}_k,\widehat{Y}_k)\big] \tag{3.17}$$

$$= \mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big] - \mathbb{E}\big[V_{k+1} \mid (X_k,Y_k)\big] \tag{3.18}$$

$$+ \mathbb{E}\big[V_{k+1} \mid (X_k,Y_k)\big] - \mathbb{E}\big[V_{k+1} \mid (\widehat{X}_k,\widehat{Y}_k)\big]. \tag{3.19}$$

Once again, both terms can be upper-bounded.

- As for the first term in the right hand side of (3.18), notice that using that

$$\mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big] - \mathbb{E}\big[V_{k+1} \mid (X_k, Y_k)\big] \\ = Pv_{k+1}(X_k, W_k^f, Y_k, W_k^d) - \mathbb{E}\big[Pv_{k+1}(X_k, W_k^f, Y_k, W_k^d) \mid (X_k, Y_k)\big]$$

where we used that $\sigma(X_k, Y_k) \subset \sigma(X_k, W_k^f, Y_k, W_k^d)$. Then, using the fact that the conditional expectation is the best quadratic approximation, we have

$$\begin{aligned}
&\big\| \mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big] - \mathbb{E}\big[V_{k+1} \mid (X_k, Y_k)\big]\big\|_2^2 \\
&\quad\leqslant \big\|Pv_{k+1}(X_k, W_k^f, Y_k, W_k^d) - Pv_{k+1}\big(X_k, \mathbb{E}[W_k^f \mid (X_k, Y_k)], Y_k, \mathbb{E}[W_k^d \mid (X_k, Y_k)]\big)\big\|_2^2 \\
&\quad\leqslant \Big\|\Big(\bar{B}_k|W_k^f - \mathbb{E}[W_k^f \mid (X_k, Y_k)]| + \bar{D}_k|W_k^d - \mathbb{E}[W_k^d \mid (X_k, Y_k)]|\Big) \\
&\qquad\qquad \times \mathrm{e}^{|Y_k| + b_k|W_k^d| \vee |\mathbb{E}[W_k^d|(X_k,Y_k)]|}\Big\|_2^2
\end{aligned}$$

where $\bar{B}_k$, $\bar{D}_k$ and $b_k = \sigma_d \mathbf{h}(n-k)$ are given in Lemma 2.3.
Hölder's inequality with $p, q \geqslant 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$ yields

$$\begin{aligned}
&\big\| \mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big] - \mathbb{E}\big[V_{k+1} \mid (X_k, Y_k)\big]\big\|_2^2 \\
&\quad\leqslant 2\widetilde{K}_k^q(b_k)\Big(\bar{B}_k^2\big\|W_k^f - \mathbb{E}[W_k^f \mid (X_k, Y_k)]\big\|_{2p}^2 + \bar{D}_k^2\big\|W_k^d - \mathbb{E}\big[W_k^d \mid (X_k, Y_k)\big]\big\|_{2p}^2\Big)
\end{aligned}$$

with the notation $\widetilde{K}_k^q(b) = \big\|e^{|Y_k| + b|W_k^d| \vee |\mathbb{E}[W_k^d|(X_k,Y_k)]|}\big\|_{2q}^2$.

- For the second term in the right hand side of (3.17), we define

$$\widetilde{v}_k(x, y) = \mathbb{E}\big[V_{k+1} \mid (X_k, Y_k) = (x, y)\big] = \mathbb{E}\big[Pv_{k+1}(x, W_k^f, y, W_k^d)\big].$$

On the other hand, note again that $(\widehat{X}_k, \widehat{Y}_k)$ being a function of (X_k, Y_k), the tower property of conditional expectation yieds $\mathbb{E}\big[V_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\big] = \mathbb{E}\big[\,\mathbb{E}\big[V_{k+1} \mid (X_k, Y_k)\big] \mid (\widehat{X}_k, \widehat{Y}_k)\big]$ combining these two identities with the property of best quadratic approximation of conditional expectation yields

$$\begin{aligned}
\big\| \mathbb{E}\big[V_{k+1} \mid (X_k, Y_k)\big] - \mathbb{E}\big[V_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\big]\big\|_2^2 &\leqslant \big\| \mathbb{E}\big[V_{k+1} \mid (X_k, Y_k)\big] - \widetilde{v}_k(\widehat{X}_k, \widehat{Y}_k)\big\|_2^2 \\
&= \big\|\widetilde{v}_k(X_k, Y_k) - \widetilde{v}_k(\widehat{X}_k, \widehat{Y}_k)\big\|_2^2.
\end{aligned}$$

Now it follows from Lemma 2.13 that $\widetilde{v}_k$ is locally-Lipschitz continuous in (x, y) namely

$$\begin{aligned}
\big\|\widetilde{v}_k(X_k, Y_k) - \widetilde{v}_k(\widehat{X}_k, \widehat{Y}_k)\big\|_2^2 &\leqslant \Big\|\Big(\bar{A}_k|X_k - \widehat{X}_k| + \bar{C}_k|Y_k - \widehat{Y}_k|\Big)e^{|Y_k| \vee |\widehat{Y}_k|}\Big\|_2^2 \\
&\leqslant 2K_k^q(0)\Big(\bar{A}_k^2\big\|X_k - \widehat{X}_k\big\|_{2p}^2 + \bar{C}_k^2\big\|Y_k - \widehat{Y}_k\big\|_{2p}^2\Big).
\end{aligned}$$

We then obtain

$$\begin{aligned}
\big\|V_k - \widehat{V}_k^{(\mathrm{NM})}\big\|_2^2 &\leqslant 2K_k^q(0)\Big((A_k^\psi)^2\big\|X_k - \widehat{X}_k\big\|_{2p}^2 + (C_k^\psi)^2\big\|Y_k - \widehat{Y}_k\big\|_{2p}^2\Big) \\
&\quad + 2\widetilde{K}_k^q(b_k)\Big(\bar{B}_k^2\big\|W_k^f - \mathbb{E}[W_k^f \mid (X_k, Y_k)]\big\|_{2p}^2 + \bar{D}_k^2\big\|W_k^d - \mathbb{E}\big[W_k^d \mid (X_k, Y_k)\big]\big\|_{2p}^2\Big) \\
&\quad + 2K_k^q(0)\Big(\bar{A}_k^2\big\|X_k - \widehat{X}_k\big\|_{2p}^2 + \bar{C}_k^2\big\|Y_k - \widehat{Y}_k\big\|_{2p}^2\Big) + \big\|V_{k+1} - \widehat{V}_{k+1}^{(\mathrm{NM})}\big\|_2^2
\end{aligned}$$

and we deduce (3.13) by induction.

Finally, applying the L^r-L^s mismatch theorem to the Gaussian distributions of (X_k, Y_k) with $r = 2$ and $s = 2p > 1$ with the quadratic optimal quantizations $\widehat{X}_k$ and $\widehat{Y}_k$ and $1 < p < 3/2$, yields

$$\limsup_{\bar{N}\to+\infty} \big\| V_k - \widehat{V}_k^{(\mathrm{NM})} \big\|_2^2 \leqslant \sum_{\ell=k}^{n-1} C_{W_\ell^f} \big\| W_\ell^f - \mathbb{E}[W_\ell^f \mid (X_\ell, Y_\ell)] \big\|_{2p}^2 + C_{W_\ell^d} \big\| W_\ell^d - \mathbb{E}[W_\ell^d \mid (X_\ell, Y_\ell)] \big\|_{2p}^2 .$$

□

Practitioner's corner. Market implied values of σ_f, σ_d and σ_S used for the numerical computations are usually of order

$$\sigma_f \approx 0.005, \qquad \sigma_d \approx 0.005, \qquad \sigma_S \approx 0.5.$$

In practice, the value of the systematic error $\mathcal{E}_n^{(\mathrm{NM})}(k)$ remaining in (3.14) is highly dependant of the value of the volatilities σ_d and σ_f. Indeed, the constants $C_{W_\ell^f}$ and $C_{W_\ell^d}$ are of order σ_f^2 and σ_d^2, respectively

$$C_{W_\ell^f} = 2\sigma_f^2 \, \mathbf{h}^2 \, \widetilde{K}_k^q(b_k) \Big(\frac{1}{\bar{\pi}_k} \sum_{\ell=k+1}^{n} \bar{\pi}_\ell \bar{A}_\ell \Big)^2, quad C_{W_\ell^d} = 2\sigma_d^2 \, \mathbf{h}^2 \, \widetilde{K}_k^q(b_k) \Big(\frac{1}{\bar{\pi}_k} \sum_{\ell=k+1}^{n} \bar{\pi}_\ell \bar{C}_\ell \Big)^2,$$

where $\bar{A}_k$ and $\bar{C}_k$ are given in (2.14) and (2.15). The following numerical experiments outline this behaviour.

4 Numerical experiments

In this section, we illustrate the theoretical results found in Section 3 regarding the pricing of Bermudan options in the 3-factor model described in Section 1. First, we detail both algorithms and how to compute the quantities that appear in them (conditional expectation, conditional probabilities, ...). Then, we test our two numerical solutions for the pricing of European options, whose price is known in closed form. European options are Bermudan options with only one date of exercise, hence when using the non-Markovian approximate we do not introduce the systematic error shown in Theorem 3.4 but pricing these kind of options is a good benchmark in order to test our methodologies. Finally, we evaluate Bermudan options and compare our two solutions, the Markovian and the non-Markovian approximation.

We have to keep in mind that the computation times are crucial because these pricers are only a small block in the complex computation of xVA's. Indeed, they will be called hundreds of thousands of times each time these risks measures are needed.

All the numerical tests have been carried out in C++ on a laptop with a 2,4 GHz 8-Core Intel Core i9 CPU. The computations of the transition probabilities and the computations of the conditional expectations are parallelized on the CPU.

Remark. The computation times given below measure the time needed for loading the precomputed optimal grids from files, rescaling the optimal quantizers in order to get the right variance, computing the conditional probabilities (the part that demands the most in term of computing power) and finally computing the expectations for the pricing. One has to keep in mind that the complexity is linear in function of n, the number of exercise dates. Indeed, if we double the number of exercise dates, we double the number of conditional probability matrices and expectations to compute.

Characterization of the Quantization Tree. In what follows, we describe the choice of parameters we made when building the quantization tree: the time discretisation and the size of each grid at each time.

- The time discretisation is an easy choice because it is decided by the characteristics of the financial product. Indeed, we take only one date (and today's date) in the tree if we want to evaluate European options and if we want to evaluate Bermudan options we take as many discretisation dates (plus today's date) in the tree as there are exercise dates in the description of the product.
- Then, we have to decide the size of each grid at each date in the tree. In our case, we consider grids of same size at each date hence $N_k = N$, $k = 1\ldots, n$ and then we take $N^X = 10N^Y$ for both trees. This choice seems to be reasonable because the risk factor X_k is prominent, due to the value of σ_S compare to σ_d. Now, in the Markovian case, we take $N^X = 4N^{W_f}$ and $N^Y = 4N^{W_d}$, indeed the two Brownian Motions are important only when we compute the conditional expectation but not when we want to evaluate the payoffs, hence we want to give as much as possible of the budget N to N^X and N^Y.

The algorithm: Markovian Case. Due to the dimension of the problem (4 in this case), we cannot compute the probabilities defined in (3.3) using deterministic methods, hence one has to simulate trajectories of the processes in order to evaluate them. We refer the reader to [BPP05, BP03, PPP04] for details on the methodology.

A way to reduce the complexity of the problem is to approximate these probabilities by $\tilde{\pi}_{i,j}^{(\mathrm{M}),k}$, where the conditional part $\widehat{E}_k^{(\mathrm{M})} = \big\{\big(\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d\big) = \big(x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k\big)\big\}$ is replaced by $E_k = \big\{(X_k, W_k^f, Y_k, W_k^d) = \big(x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k\big)\big\}$, yielding

$$\begin{aligned}\tilde{\pi}_{i,j}^{(\mathrm{M}),k} = \mathbb{P}\Big(\big(\widehat{X}_{k+1}, \widehat{W}_{k+1}^f, \widehat{Y}_{k+1}, \widehat{W}_{k+1}^d\big) = \big(x_{j_1}^{k+1}, u_{j_2}^{k+1}, y_{j_3}^{k+1}, v_{j_4}^{k+1}\big) \\ \mid (X_k, W_k^f, Y_k, W_k^d) = \big(x_{i_1}^k, y_{i_2}^k, u_{i_3}^k, v_{i_4}^k\big)\Big).\end{aligned} \tag{4.1}$$

The reason for replacing $\widehat{E}_k^{(\mathrm{M})}$ by E_k is explained in the next paragraph dealing with the Non-Markovian case with lighter notations (see Equation (4.2) and (4.3)). Although, these probabilities are easier to calculate, one still has to devise a Monte Carlo simulation in order to evaluate them. This simplification will be useful later in the uncorrelated case.

Following these remarks, the QBDPP in the Markovian case (3.2) rewrites as

$$\begin{cases}\widehat{v}_n\big(x_{i_1}^n, u_{i_2}^n, y_{i_3}^n, v_{i_4}^n\big) = h_n\big(x_{i_1}^n, y_{i_3}^n\big), \\ \widehat{v}_k\big(x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k\big) = \max\Big(h_k\big(x_{i_1}^k, y_{i_3}^k\big), \displaystyle\sum_{j_1,j_2,j_3,j_4} \tilde{\pi}_{i,j}^{(\mathrm{M}),k}\, \widehat{v}_{k+1}\big(x_{j_1}^{k+1}, u_{j_2}^{k+1}, y_{j_3}^{k+1}, v_{j_4}^{k+1}\big)\Big).\end{cases}$$

The algorithm: Non-Markovian case. The transition probabilities (3.12) can be computed by numerical integration. Denoting $C_i(\Gamma_x^k) = (x_{i-1/2}^k, x_{i+1/2}^k)$ we have

$$\begin{aligned}\pi_{i,j}^{(\mathrm{NM}),k} &= \mathbb{P}\Big(\big(\widehat{X}_{k+1}, \widehat{Y}_{k+1}\big) = \big(x_{j_1}^{k+1}, y_{j_2}^{k+1}\big) \mid \big(\widehat{X}_k, \widehat{Y}_k\big) = \big(x_{i_1}^k, y_{i_2}^k\big)\Big) \\ &= \mathbb{P}\Big(\big(\widehat{X}_{k+1}, \widehat{Y}_{k+1}\big) = \big(x_{j_1}^{k+1}, y_{j_2}^{k+1}\big) \mid X_k \in C_{i_1}(\Gamma_x^k), Y_k \in C_{i_2}(\Gamma_y^k)\Big) \\ &= \frac{\mathbb{P}\Big(X_{k+1} \in C_{j_1}(\Gamma_x^{k+1}), Y_{k+1} \in C_{j_2}(\Gamma_y^{k+1}), X_k \in C_{i_1}(\Gamma_x^k), Y_k \in C_{i_2}(\Gamma_y^k)\Big)}{\mathbb{P}\Big(X_k \in C_{i_1}(\Gamma_x^k), Y_k \in C_{i_2}(\Gamma_y^k)\Big)}.\end{aligned} \tag{4.2}$$

The vectors $(X_{k+1}, Y_{k+1}, X_k, Y_k)$ and (X_k, Y_k) are Gaussian vectors and their covariance matrices are known. It is therefore possible to compute the probabilities in Equation (4.2) by numerical integrations (in dimension 4 for the numerator and in dimension 2 for the denominator). These numerical integrations could be too time consuming, hence once again, we approximate these probabilities by $\widetilde{\pi}_{i,j}^{(\text{NM}),k}$, where the conditional part $\{(\widehat{X}_k, \widehat{Y}_k) = (x_{i_1}^k, y_{i_2}^k)\}$ is replaced by $\{(X_k, Y_k) = (x_{i_1}^k, y_{i_2}^k)\}$, yielding

$$\widetilde{\pi}_{i,j}^{(\text{NM}),k} = \mathbb{P}\Big((\widehat{X}_{k+1}, \widehat{Y}_{k+1}) = (x_{j_1}^{k+1}, y_{j_2}^{k+1}) \mid (X_k, Y_k) = (x_{i_1}^k, y_{i_2}^k)\Big). \tag{4.3}$$

From the definition of an optimal quantizer and Equation (2.4), this probability can be rewritten as the probability that a correlated bivariate normal distribution lies in a rectangular domain [2]

$$\begin{aligned}
\widetilde{\pi}_{i,j}^{(\text{NM}),k} &= \mathbb{P}\Big(\widehat{X}_{k+1} = x_{j_1}^{k+1}, \widehat{Y}_{k+1} = y_{j_2}^{k+1} \mid X_k = x_{i_1}^k, Y_k = y_{i_2}^k\Big)\\
&= \mathbb{P}\Big(X_{k+1} \in (x_{j_1-1/2}^{k+1}, x_{j_1+1/2}^{k+1}), Y_{k+1} \in (y_{j_2-1/2}^{k+1}, y_{j_2+1/2}^{k+1}) \mid X_k = x_{i_1}^k, Y_k = y_{i_2}^k\Big)\\
&= \mathbb{P}\Big(x_{i_1}^k + \sigma_f \mathbf{h} W_k^f + G_{k+1}^1 \in (x_{j_1-1/2}^{k+1}, x_{j_1+1/2}^{k+1}), y_{i_2}^k - \sigma_d \mathbf{h} W_k^d + G_{k+1}^3 \in (y_{j_2-1/2}^{k+1}, y_{j_2+1/2}^{k+1})\Big)\\
&= \mathbb{P}\Big(Z^1 \in (x_{j_1-1/2}^{k+1} - x_{i_1}^k, x_{j_1+1/2}^{k+1} - x_{i_1}^k), Z^2 \in (y_{j_2-1/2}^{k+1} - y_{i_2}^k, y_{j_2+1/2}^{k+1} - y_{i_2}^k)\Big)
\end{aligned} \tag{4.5}$$

where

$$\begin{pmatrix} Z^1 \\ Z^2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{Z^1}^2 & \rho_{Z^1,Z^2}\sigma_{Z^1}\sigma_{Z^2} \\ \rho_{Z^1,Z^2}\sigma_{Z^1}\sigma_{Z^2} & \sigma_{Z^2}^2 \end{pmatrix}\right)$$

with $\sigma_{Z^1}^2 = \mathbb{V}\mathrm{ar}(\sigma_f \mathbf{h} W_k^f + G_{k+1}^1)$, $\sigma_{Z^2}^2 = \mathbb{V}\mathrm{ar}(-\sigma_d \mathbf{h} W_k^d + G_{k+1}^3)$ and $\rho_{Z^1,Z^2} = \mathbb{C}\mathrm{orr}(\sigma_f \mathbf{h} W_k^f + G_{k+1}^1, -\sigma_d \mathbf{h} W_k^d + G_{k+1}^3)$.

Now, following the above footnote, if (U, V) is a two-dimensional correlated and standardised normal distribution with correlation ρ and cumulative distribution function (CDF) given by $F_{U,V}^{\rho}(u, v) = \mathbb{P}(U \leqslant u, V \leqslant v)$, $u, v \in \mathbb{R}$, one has for (u_1, v_1), (u_2, v_2), $u_1 \leqslant u_2$, $v_1 \leqslant v_2$,

$$\mathbb{P}\big(U \in (u_1, u_2), V \in (v_1, v_2)\big) = F_{U,V}^{\rho}(u_2, v_2) - F_{U,V}^{\rho}(u_1, v_2) - F_{U,V}^{\rho}(u_2, v_1) + F_{U,V}^{\rho}(u_1, v_1). \tag{4.6}$$

This remark applied to (4.5) with $U = Z^1/\sigma_{Z^1}$ and $V = Z^2/\sigma_{Z^2}$ will allow us to reduce drastically the computation time induced by the evaluation of the conditional probabilities and so, of the conditional expectations.

[2] The advantage of expressing (4.5) as the probability that a bivariate Gaussian vector lies in a rectangular domain is that it can be rewritten as a linear combination of bivariate cumulative distribution functions. Indeed, let (U, V) a two-dimensional correlated and standardized normal distribution with correlation ρ and cumulative distribution function (CDF) given by $F_{U,V}^{\rho}(u, v) = \mathbb{P}(U \leqslant u, V \leqslant v)$. Fast and efficient numerical implementation of such function exists (for example, a C++ implementation of the upper right tail of a correlated bivariate normal distribution can be found in John Burkardt's website, see [Bur12], which is based on the work of [Don73] and [Owe58].
We are interested in the computation of probabilities of the form

$$\mathbb{P}\big(U \in (u_1, u_2), V \in (v_1, v_2)\big). \tag{4.4}$$

This probability is represented graphically as the integral of the two-dimensional density over the rectangular domain in grey in Figure 2.

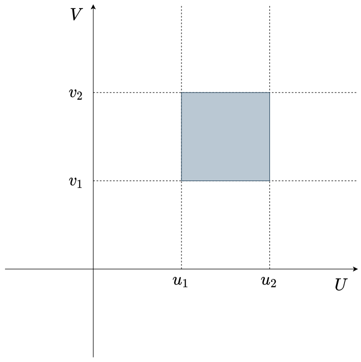


Figure 2: Domain of integration for probabilities of correlated two-dimensional Gaussian random vector.

Now, going back to our problem, the QBDPP in the non-Markovian case rewrites (3.11)

$$\begin{cases} \widehat{v}_n\big(x^n_{i_1}, y^n_{i_2}\big) = h_n\big(x^n_{i_1}, y^n_{i_2}\big), \qquad 1 \leqslant i_1 \leqslant N^X_n, \qquad 1 \leqslant i_2 \leqslant N^Y_n, \\ \widehat{v}_k\big(x^k_{i_1}, y^k_{i_2}\big) = \max\Big(h_k\big(x^k_{i_1}, y^k_{i_2}\big), \sum_{j_1,j_2} \pi^{(\mathrm{NM}),k}_{i,j}\, \widehat{v}_{k+1}\big(x^{k+1}_{j_1}, y^{k+1}_{j_2}\big)\Big). \end{cases}$$

In order to test numerically the two methods, we will evaluate PRDC European and Bermudan options with maturities $2Y$, $5Y$, $10Y$ and $20Y$. We describe below the market and products parameters we consider. The volatilities of the domestic and the foreign interest rates are not detailed below because we investigate the behaviour of the methods with respect to σ_d and σ_f.

$P^d(0,t)$	$\exp(-r_d t)$	r_d	0.015
$P^f(0,t)$	$\exp(-r_f t)$	r_f	0.01
S_0	88.17	σ_S	0.5

Table 1: *Market parameters. The so-called 'zero correlation case' means* $\rho_{Sd} = \rho_{Sf} = \rho_{df} = 0$*, otherwise the correlations are fixed at* $\rho_{Sf} = -0.0272$*,* $\rho_{Sd} = 0.1574$*, and* $\rho_{df} = 0.6558$.

$\forall\, k \in 1,\dots,n,\quad C_d(t_k)$	15%	$\forall\, k \in 1,\dots,n,\quad C_f(t_k)$	18.9%
$\forall\, k \in 1,\dots,n,\quad \mathrm{Cap}(t_k)$	5.55%	$\forall\, k \in 1,\dots,n,\quad \mathrm{Floor}(t_k)$	0%
Exercise date (EU): t_n	T	Exercise dates (US): t_k	Tk/n

Table 2: *Product description.*

Remark. When the correlations ρ_{df} and ρ_{Sd} are equal to zero, the numerical computation of probabilities $\widetilde{\pi}^{(\mathrm{M}),k}_{i,j}$ and $\widetilde{\pi}^{(\mathrm{NM}),k}_{i,j}$ can be accelerated. Indeed, in the Markovian case, Equation (4.1) can be rewritten as

$$\begin{aligned} \widetilde{\pi}^{(\mathrm{M}),k}_{i,j} = \mathbb{P}\Big((\widehat{X}_{k+1}, \widehat{W}^f_{k+1}) = \big(x^{k+1}_{j_1}, u^{k+1}_{j_2}\big) \mid (X_k, W^f_k) = \big(x^k_{i_1}, u^k_{i_2}\big)\Big) \\ \times\, \mathbb{P}\Big((\widehat{Y}_{k+1}, \widehat{W}^d_{k+1}) = \big(y^{k+1}_{j_3}, v^{k+1}_{j_4}\big) \mid (Y_k, W^d_k) = \big(y^k_{i_3}, v^k_{i_4}\big)\Big). \end{aligned}$$

In that case, we can use the CDF of a correlated bivariate normal distribution, as detailed above for the non-Markovian case in (4.6), for computing these probabilities in a very effective and faster way rather than performing a Monte Carlo simulation.

In the non-Markovian case, Equation (4.5) can be rewritten as

$$\begin{aligned}
\widetilde{\pi}_{i,j}^{(\mathrm{NM}),k} &= \mathbb{P}\left(Z^1 \in \left(x_{j_1-1/2}^{k+1} - x_{i_1}^k, x_{j_1+1/2}^{k+1} - x_{i_1}^k\right)\right)\mathbb{P}\left(Z^2 \in \left(y_{j_2-1/2}^{k+1} - y_{i_2}^k, y_{j_2+1/2}^{k+1} - y_{i_2}^k\right)\right)\\
&= \left(F_Z\left(\frac{x_{j_1+1/2}^{k+1} - x_{i_1}^k}{\sigma_{Z^1}}\right) - F_Z\left(\frac{x_{j_1-1/2}^{k+1} - x_{i_1}^k}{\sigma_{Z^1}}\right)\right)\left(F_Z\left(\frac{y_{j_2+1/2}^{k+1} - y_{i_2}^k}{\sigma_{Z^2}}\right) - F_Z\left(\frac{y_{j_2-1/2}^{k+1} - y_{i_2}^k}{\sigma_{Z^2}}\right)\right)
\end{aligned}$$

where $F_Z(\cdot)$ is the CDF of a one-dimensional normal distribution, σ_{Z^1} is the standard deviation of Z^1 and σ_{Z^2} is the standard deviation of Z^2. This remark allows us to drastically reduce the computation time of the conditional probabilities in the case of zero correlations.

4.1 European Option

In this short section, we look at the pricing of European options. In this simple case, the payoff $h_T(X_T, Y_T)$ depends only on the maturity date T, and the random variables (X_T, Y_T) (see (2.3)). We the consider the algorithm of section 3.3 with two dates $t_0 = 0$ and $t_n = T$ ($n = 1$) in the recursion (3.11) (see also (4.5)). Note that in this European pricing problem, there is no systematic error induced by the non-markovianity of the couple (X_n, Y_n). In other words, the error term $\mathcal{E}_n^{(\mathrm{NM})}(n)$ is equals to zero in Theorem 3.4.

In the case of the European options, we have a closed-form formula for the price of (2.1). The benchmark price is computed using the rewriting of (2.1) as a sum of Calls: at a time t_k, the payoff can be expressed as

$$\begin{aligned}
\psi_{t_k}(S_{t_k}) &= \min\left(\max\left(\frac{C_f(t_k)}{S_0}S_{t_k} - C_d(t_k), \mathrm{Floor}(t_k)\right), \mathrm{Cap}(t_k)\right)\\
&= \mathrm{Floor}(t_k) - a_k(S_{t_k} - K_k^1)_+ + a_k(S_{t_k} - K_k^2)_+
\end{aligned}$$

with $a_k = \dfrac{C_f(t_k)}{S_0}$, $K_k^1 = \dfrac{\mathrm{Cap}(t_k) + C_d(t_k)}{C_f(t_k)} \times S_0$ and $K_k^2 = \dfrac{\mathrm{Floor}(t_k) + C_d(t_k)}{C_f(t_k)} \times S_0$ and the closed-form formula for the price of a Call is detailed in Appendix B. The exact prices are summarised in Table 3.

	Zero correlation case		Correlated case	
T \ σ	50 bps	500 bps	50 bps	500 bps
2Y	2.171945242	2.159404007	2.173803852	2.185536786
5Y	1.630435483	1.539295559	1.636518082	1.652226813
10Y	1.127330259	0.8013151892	1.141944391	1.103531914
20Y	0.5953823852	0.07313312587	0.6262982227	0.3764765391

Table 3: *Prices given by closed-form formula of European options ($\sigma_d = \sigma_f = \sigma$).*

We compute the relative error between the price computed by the 2d–quantization with N points denoted $\widehat{V}_0^N$ and the exact price V_0 defined by $(\widehat{V}_0^N - V_0)/V_0$. The size $N = N^X \times N^Y$ of the product quantizer, and the associated computation times, needed for a 1 bps relative error are summarised in Table 4.

T \ σ	Zero correlation case 50 bps	Zero correlation case 500 bps	Correlated case 50 bps	Correlated case 500 bps
2Y	1 ms (32000)	4 ms (32000)	71 ms (64000)	34 ms (32000)
5Y	4 ms (32000)	6 ms (32000)	31 ms (32000)	31 ms (32000)
10Y	4 ms (32000)	3 ms (32000)	32 ms (32000)	139 ms (128000)
20Y	2 ms (32000)	2 ms (32000)	54 ms (32000)	2147 ms (2048000)

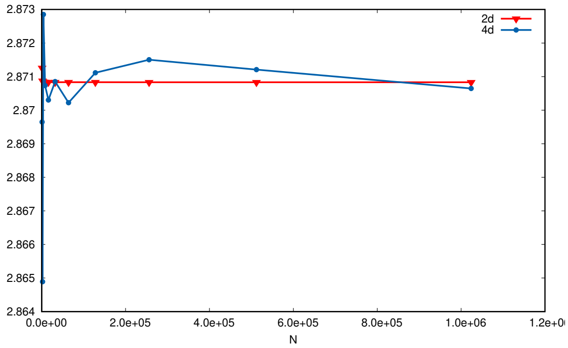
Table 4: *Times in milliseconds needed for reaching a* 1 bps *precision for European options pricing, in parenthesis, the size N of the grid ($\sigma_d = \sigma_f = \sigma$).*

It is noteworthy that in the zero correlation case a relative error of 1 bps is very quickly reached, even for high values of σ_d and σ_f. Indeed in this special case, it suffices to take $N = 32\,000$ to achieve a precision of 1 bps. In the correlated case with values $\sigma_d = \sigma_f = 50$ bps (order of magnitude of market values), the pricing by quantization is also very efficient.

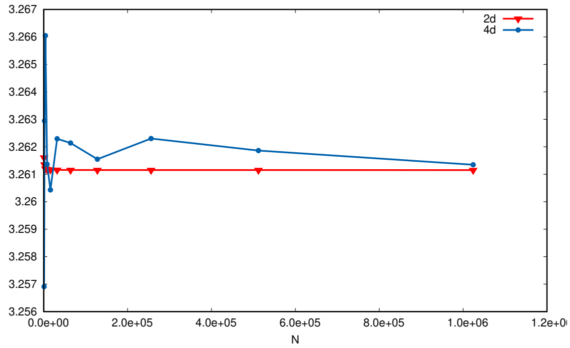
4.2 Bermudan option

Now, we compare the asymptotic behaviour of both approaches (Markovian case and Non-Markovian case) when pricing Bermudan PRDC options. The following figures represent the price and the rescaled difference of the prices given by the two approaches as a function of N, which is the size of the product quantizer at each date (in two dimensions: $N = N^X \times N^Y$ and in four dimensions $N = N^X \times N^{W^f} \times N^Y \times N^{W^d}$). The financial products we consider are yearly exercisable Bermudan options with different values for the maturity date (2 years, 5 years, 10 years and 20 years) and the domestic/foreign volatilities (50 bps and 500 bps). The Bermudan options we consider can be exercised every year, so the number n of iterations in the dynamic programming satisfies $n = M + 1$ where M is the maturity date, $M \in \{2, 5, 10, 20\}$.

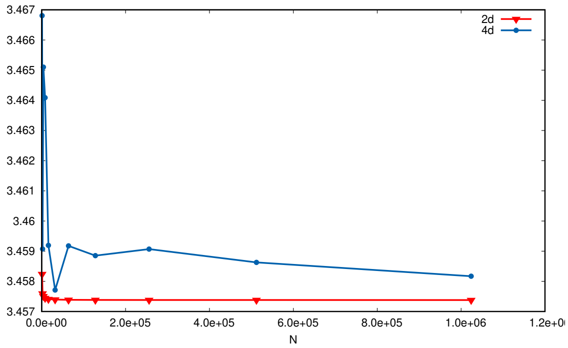
When using domestic and foreign volatilities close to market values, we observe numerically that the non-Markovian method converges a lot faster than the Markovian one for a given complexity. However both methods do not converge to the same value (see Figures 3a, 3b, 3c), which is consistent with the results we found in Theorems 3.3 and 3.4. On each plot of Figure 3, we observe that the red curve (non-Markovian method in 2d) stabilizes very quickly while the blue curve (Markovian method in 4d) takes time to stabilize. From a large N, the blue curve decreases very slowly towards an asymptotic value which represents the price of the product. The convergence is fast in the non-Markovian case because the dimension of the problem is 2, and slower in the Markovian case because the dimension is 4.



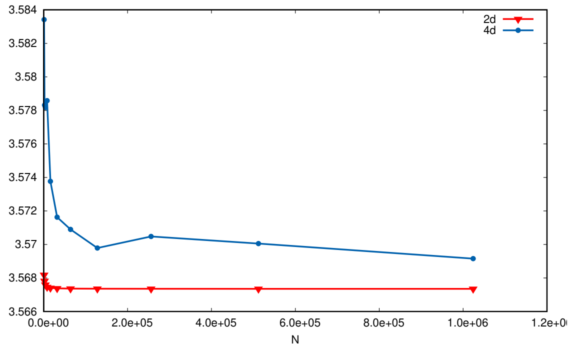
(a) *2Y*



(b) *5Y*



(c) *10Y*



(d) *20Y*

Figure 3: $\sigma_d = \sigma_f = 50\,\text{bps}$ – *Prices with the two methods for 2Y, 5Y, 10Y and 20Y yearly exercisable Bermudan options (zero correlation case).*

In Figure 4, we display the relative differences $|\widehat{V}_0^{(\mathrm{NM})} - \widehat{V}_0^{(\mathrm{M})}|/V_0^{(\mathrm{M})}$ (the rescaled absolute errors between prices given by the non-Markovian method and the Markovian one) between the two methods, as a function of N. We then approximate the so-called relative systematic error (3.14) by $|\widehat{V}_0^{(\mathrm{NM})} - \widehat{V}_0^{(\mathrm{M})}|/V_0^{(\mathrm{M})}$ for N large enough. This error, for different maturity dates, can be read in Figure 4 for $N = 1\,000\,000$. Note that this relative systematic error is negligible for theses values of $\sigma_f = \sigma_d = 50\,\text{bps}$. It is at most $5\,\text{bps}$ for a 20-year annual Bermudan option.

Hence, in the case of standard market of values for the volatilities, one should prefer the non-Markovian methodology when considering to evaluate Bermudan options.

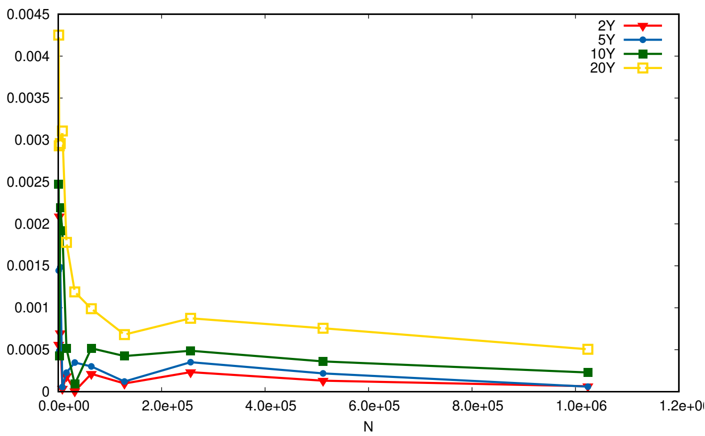
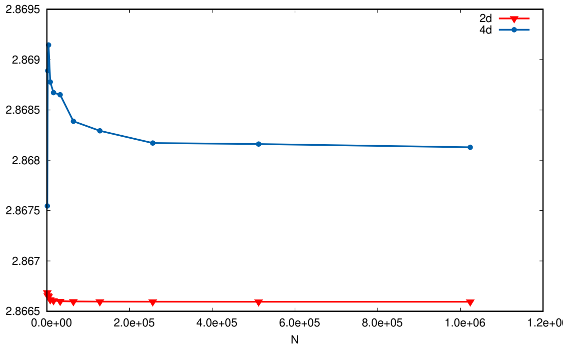
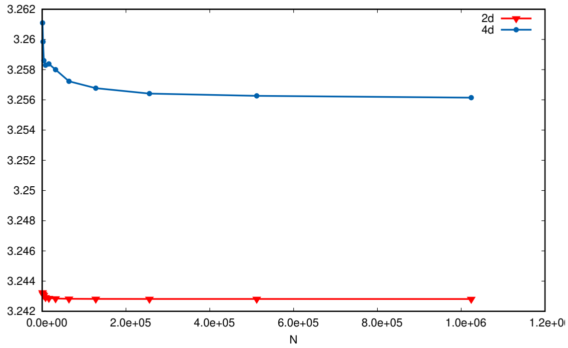


Figure 4: $\sigma_d = \sigma_f = 50\,\text{bps}$ – *Relative differences between the two methods for 2Y, 5Y, 10Y and 20Y yearly exercisable Bermudan options (zero correlation case).*

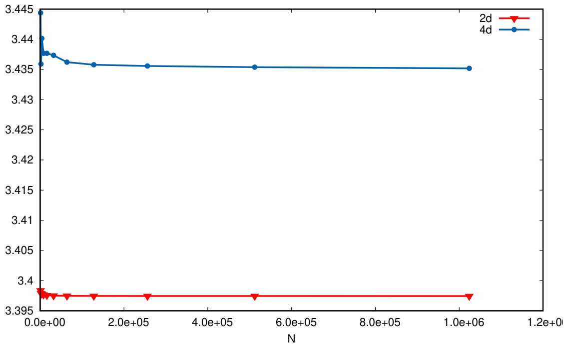
When we consider higher values the volatilities, $\sigma_d = \sigma_f = 500\,\text{bps}$, as expected the non-Markovian methodology produces a systematic error bigger than the case where $\sigma_d = \sigma_f = 50\,\text{bps}$ (see Figures 5a, 5b, 5c and 6). The relative systematic errors between the two methods are reasonable: less than 0.1% for expiry 2 years, 0.4% for 5 years, around 1.1% for 10 years and 1.6% for 20 years. Again, one has to bear in mind that these volatilities are not at the scale of those observed on the market. We chose to display these numerical results in order to illustrate the limitations of the non-Markovian methodology and to push to its limits.



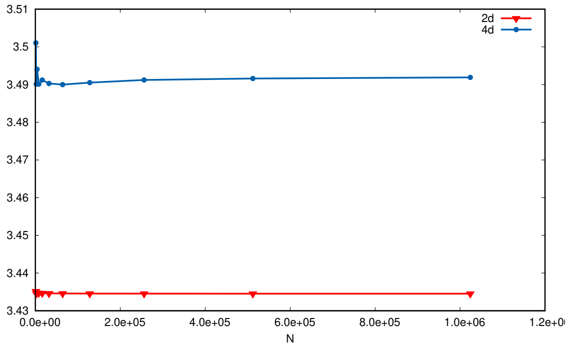
(a) *2Y*



(b) *5Y*



(c) *10Y*



(d) *20Y*

Figure 5: $\sigma_d = \sigma_f = 500\,\text{bps}$ – *Prices with the two methods for 2Y, 5Y, 10Y and 20Y yearly exercisable Bermudan options (zero correlation case).*

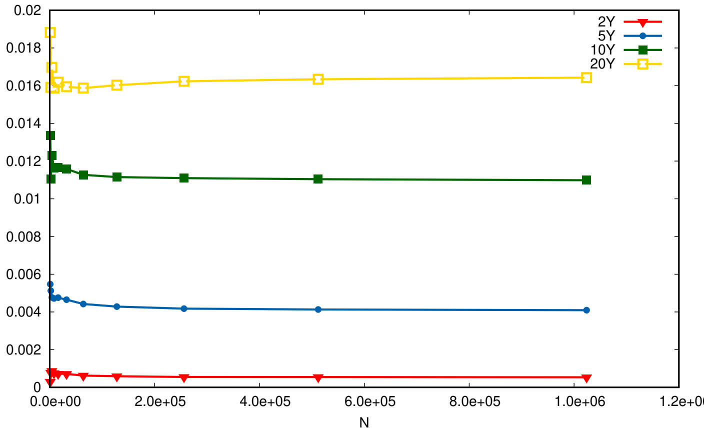


Figure 6: $\sigma_d = \sigma_f = 500\,\text{bps}$ – *Relative differences between the two methods for 2Y, 5Y, 10Y and 20Y yearly exercisable Bermudan options (zero correlation case).*

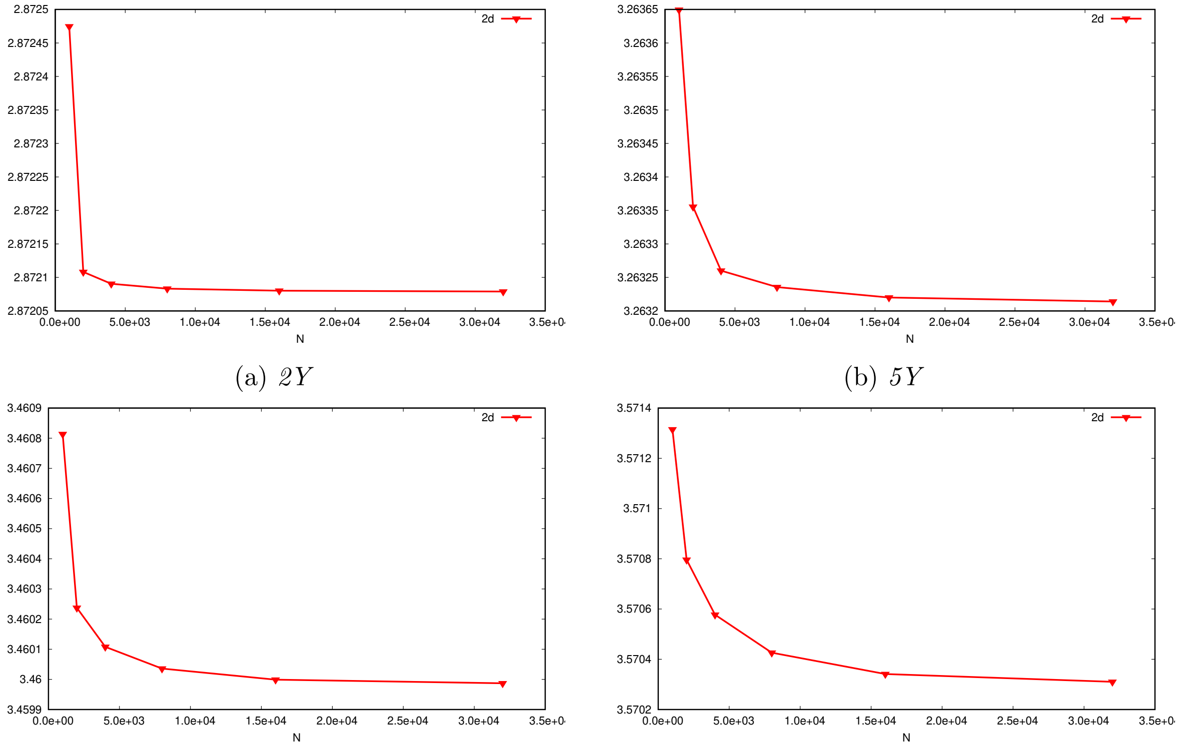
We are now interested in the speed of convergence of the two algorithms. For that we consider the number N which must be chosen so that the error is in a $5\,\text{bps}$ band around the asymptotic value ($\widehat{V}_0^{(\text{M}),\infty}$ and $\widehat{V}_0^{(\text{NM}),\infty}$ respectively) approached with $N = 1\,000\,000$. More precisely in Table 5, we reference the time needed for reaching the asymptotic price with a $5\,\text{bps}$ precision for each algorithm (the size N of the grid is $\min\{N, \forall M \geqslant N, |\widehat{V}_0^{(\text{M}),M} - \widehat{V}_0^{(\text{M}),\infty}|/\widehat{V}_0^{(\text{M}),\infty} < 5\,\text{bps}\}$ and $\min\{N, \forall M \geqslant N, |\widehat{V}_0^{(\text{NM}),M} - \widehat{V}_0^{(\text{NM}),\infty}|/\widehat{V}_0^{(\text{NM}),\infty} < 5\,\text{bps}\}$ respectively). The non-Markovian method converges very quickly and attains better precision than a relative precision of $5\,\text{bps}$ in a few milliseconds, at most 7 ms. This is a very well feature of this algorithm. The Markovian one, which is in dimension 4, converges slowly.

	Non-Markovian – 2d		Markovian – 4d	
T \ σ	$50\,\text{bps}$	$500\,\text{bps}$	$50\,\text{bps}$	$500\,\text{bps}$
2Y	1 ms (1000)	1 ms (1000)	25 ms (8000)	4 ms (1000)
5Y	3 ms (1000)	4 ms (1000)	98 ms (8000)	1903 ms (64000)
10Y	7 ms (1000)	7 ms (1000)	468 ms (16000)	3850 ms (64000)
20Y	17 ms (1000)	15 ms (1000)	8076 ms (64000)	28307 ms (128000)

Table 5: *Times in milliseconds needed for reaching the asymptotic price of the algorithm with a* $5\,\text{bps}$ *relative precision (zero correlation case). In parenthesis, the size N of the grid at each time step.*

In the correlated case, we choose to show only the asymptotic behaviour of the non-Markovian method (2d). Indeed, if we want to use the Markovian approach, we need to compute the transition probabilities using a Monte Carlo simulation which is a major drawback for the method.

Figures 7a, 7b and 7c display the price given by the numerical method as a function of N and Table 6 summarises the computation time needed in order to do better than a $5\,\text{bps}$ relative precision (error relative to the asymptotic value of the algorithm $\widehat{V}_0^{(\text{NM}),\infty}$).



(a) *2Y* (b) *5Y*

(c) *10Y* (d) *20Y*

Figure 7: $\sigma_d = \sigma_f = 50\,\text{bps}$ – *Price of 2Y, 5Y, 10Y and 20Y yearly exercisable Bermudan options using the non-Markovian method (correlated case).*

	Non-Markovian – 2d
σ / T	50 bps
2Y	122 ms (1000)
5Y	553 ms (1000)
10Y	1283 ms (1000)
20Y	2870 ms (1000)

Table 6: *Times in milliseconds needed for reaching the asymptotic price of the algorithm with a* 3 bps *relative precision (correlated case). In parenthesis, the size N of the grid at each time step.*

Conclusion

We were looking for a numerical method able to produce accurate prices of Bermudan PRDC options with a 3-factor model in a very short time because the pricing of such products arises in a more complex framework: the computation of counterparty risk measures, also called xVA's. We proposed two numerical methods based on product optimal quantization with a preference for the non-Markovian one. Indeed, even if we introduce a systematic error with our approximation, the error is controlled, as long as the volatilities of the domestic and foreign interest rates stay reasonable. Moreover, the numerical tests we conducted confirmed that idea: we are able to produce prices of Bermudan options in the 3-factor model in a fast and accurate way.

Declaration of Interest

The authors report no conflicts of interest. The authors are responsible for the content and writing of the paper.

Acknowledgments

The PhD thesis of Thibaut Montes is funded by a CIFRE grand from The Independent Calculation Agent (ICA) and French ANRT.

References

[BGM97] Alan Brace, Dariusz Gatarek, and Marek Musiela. The market model of interest rate dynamics. *Mathematical finance*, 7(2):127–155, 1997.

[BMP13] Damiano Brigo, Massimo Morini, and Andrea Pallavicini. *Counterparty credit risk, collateral and funding: with pricing cases for all asset classes*, volume 478. John Wiley & Sons, 2013.

[BP03] Vlad Bally and Gilles Pagès. A quantization algorithm for solving multidimensional discrete-time optimal stopping problems. *Bernoulli*, 9(6):1003–1049, 2003.

[BPP05] Vlad Bally, Gilles Pagès, and Jacques Printems. A quantization tree method for pricing and hedging multi-dimensional american options. *Mathematical Finance*, 15(1):119–168, 2005.

[BS73] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654, 1973.

[Bur12] John Burkardt. C++ library which evaluates the upper right tail of the bivariate normal distribution, 2012.

[CAGM97] Juan Cuesta-Albertos, Alfonso Gordaliza, and Carlos Matrán. Trimmed k-means: an attempt to robustify quantizers. *The Annals of Statistics*, 25(2):553–576, 1997.

[CBB14] Stéphane Crépey, Tomasz R Bielecki, and Damiano Brigo. *Counterparty risk and funding: A tale of two puzzles*. Chapman and Hall/CRC, 2014.

[Don73] Thomas G Donnelly. Algorithm 462: Bivariate normal distribution. *Communications of the ACM*, 16(10):638, 1973.

[EKFG96] Nicole El Karoui, Antoine Frachot, and Hélyette Geman. A note on the behavior of long zero coupon rates in a no arbitrage framework. *Preprint*, 1996.

[EKMV92] Nicole El Karoui, Ravi Myneni, and Ramanarayanan Viswanathan. Arbitrage pricing and hedging of interest rate claims with state variables, theory and applications. *Preprint*, 1992.

[GEKR95] Hélyette Geman, Nicole El Karoui, and Jean-Charles Rochet. Changes of numeraire, changes of probability measure and option pricing. *Journal of Applied probability*, pages 443–458, 1995.

[GG82] Allen Gersho and Robert M Gray. Special issue on quantization. *IEEE Transactions on Information Theory*, 29, 1982.

[GL00] Siegfried Graf and Harald Luschgy. *Foundations of Quantization for Probability Distributions*. Springer-Verlag, Berlin, Heidelberg, 2000.

[GLP08] Siegfried Graf, Harald Luschgy, and Gilles Pagès. Distortion mismatch in the quantization of probability measures. *ESAIM: Probability and Statistics*, 12:127–153, 2008.

[Gre15] Jon Gregory. *The xVA Challenge: counterparty credit risk, funding, collateral and capital.* John Wiley & Sons, 2015.

[HJM92] David Heath, Robert Jarrow, and Andrew Morton. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica: Journal of the Econometric Society*, pages 77–105, 1992.

[HW93] John Hull and Alan White. One-factor interest-rate models and the valuation of interest-rate derivative securities. *Journal of financial and quantitative analysis*, 28(2):235–254, 1993.

[Kie83] John C Kieffer. Uniqueness of locally optimal quantizer for log-concave density and convex error weighting function. *IEEE Transactions on Information Theory*, 29(1):42–47, 1983.

[LMP19] Vincent Lemaire, Thibaut Montes, and Gilles Pagès. New weak error bounds and expansions for optimal quantization. *Journal of Computational and Applied Mathematics*, page 112670, 2019.

[NP14] João Pedro Vidal Nunes and Pedro Miguel Silva Prazeres. Pricing swaptions under multifactor gaussian hjm models. *Mathematical Finance*, 24(4):762–789, 2014.

[Owe58] Donald B Owen. *Tables for computing bivariate normal probabilities.* Sandia Corporation, 1958.

[Pag98] Gilles Pagès. A space quantization method for numerical integration. *Journal of computational and applied mathematics*, 89(1):1–38, 1998.

[Pag15] Gilles Pagès. Introduction to vector quantization and its applications for numerics. *ESAIM: proceedings and surveys*, 48:29–79, 2015.

[Pag18] Gilles Pagès. *Numerical Probability: An Introduction with Applications to Finance.* Springer, 2018.

[Pit05] Vladimir Piterbarg. A multi-currency model with fx volatility skew. *SSRN Electronic Journal*, 2005.

[PPP04] Gilles Pagès, Huyên Pham, and Jacques Printems. *Optimal Quantization Methods and Applications to Numerical Problems in Finance*, pages 253–297. Birkhäuser Boston, 2004.

[PS18] Gilles Pagès and Abass Sagna. Improved error bounds for quantization based numerical schemes for bsde and nonlinear filtering. *Stochastic Processes and their Applications*, 128(3):847–883, 2018.

[She97] William Fleetwood Sheppard. On the calculation of the most probable values of frequency-constants, for data arranged according to equidistant division of a scale. *Proceedings of the London Mathematical Society*, 1(1):353–380, 1897.

[Shr04] Steven E Shreve. *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer Science & Business Media, 2004.

[Vas77] Oldrich Vasicek. An equilibrium characterization of the term structure. *Journal of financial economics*, 5(2):177–188, 1977.

[Wys17] Uwe Wystup. *FX options and structured products.* John Wiley & Sons, 2017.

Appendices

A W^f is a Brownian motion under the domestic risk-neutral measure

Let $(\widetilde{W}^f)$ a $\widetilde{\mathbb{P}}$-Brownian motion. In this section, we show that the process W^f defined by

$$dW_s^f = d\widetilde{W}_s^f + \rho_{Sf}\sigma_S ds \tag{A.1}$$

is a $\mathbb{P}$-Brownian motion.

First, we define the following change of numéraire, where $\widetilde{\mathbb{P}}$ is the foreign risk-neutral probability and $\mathbb{P}$ is the domestic risk-neutral probability,r

$$\begin{aligned} d\widetilde{\mathbb{P}} &= \frac{S_T}{S_0}\exp\left(-\int_0^T r_s^d ds\right)\exp\left(\int_0^T r_s^f ds\right)d\,\mathbb{P} \\ &= \exp\left(\sigma_S W_T^S - \frac{\sigma_S^2}{2}T\right)d\,\mathbb{P} \end{aligned}$$

or equivalently

$$\begin{aligned} d\,\mathbb{P} &= \exp\left(-\sigma_S W_T^S + \frac{\sigma_S^2}{2}T\right)d\widetilde{\mathbb{P}} \\ &= \exp\left(-\sigma_S(W_T^S - \sigma_S T) - \frac{\sigma_S^2}{2}T\right)d\widetilde{\mathbb{P}} \\ &= \exp\left(-\sigma_S \widetilde{W}_T^S - \frac{\sigma_S^2}{2}T\right)d\widetilde{\mathbb{P}} \end{aligned} \tag{A.2}$$

where $\widetilde{W}^S$ is a $\widetilde{\mathbb{P}}$-Brownian motion defined by $d\widetilde{W}_t^S = dW_t^S - \sigma_S dt$. More details concerning the definition of the foreign risk-neutral probability can be found in the Chapter 9 of [Shr04].

Now, we are looking for $q \in \mathbb{R}$ such that $dW_s^f = d\widetilde{W}_s^f - qdt$ is a $\mathbb{P}$-Brownian motion. Let $\lambda \in \mathbb{R}$ and $\forall t > s$

$$\begin{aligned} \mathbb{E}\left[\mathrm{e}^{\lambda\left((\widetilde{W}_t^f - qt) - (\widetilde{W}_s^f - qs)\right)} \mid \mathcal{F}_s\right] &= \widetilde{\mathbb{E}}\left[\mathrm{e}^{\lambda\left((\widetilde{W}_t^f - qt) - (\widetilde{W}_s^f - qs)\right) - \sigma_S(\widetilde{W}_T^S - \widetilde{W}_s^S) - \frac{\sigma_S^2}{2}(T-s)} \mid \mathcal{F}_s\right] \\ &= \widetilde{\mathbb{E}}\left[\mathrm{e}^{\lambda\left((\widetilde{W}_t^f - qt) - (\widetilde{W}_s^f - qs)\right) - \sigma_S(\widetilde{W}_t^S - \widetilde{W}_s^S) - \frac{\sigma_S^2}{2}(t-s)} \mid \mathcal{F}_s\right] \\ &= \mathrm{e}^{-\lambda q(t-s) - \frac{\sigma_S^2}{2}(t-s)}\,\widetilde{\mathbb{E}}\left[\mathrm{e}^{\lambda(\widetilde{W}_t^f - \widetilde{W}_s^f) - \sigma_S(\widetilde{W}_t^S - \widetilde{W}_s^S)} \mid \mathcal{F}_s\right] \\ &= \mathrm{e}^{-\lambda q(t-s) - \frac{\sigma_S^2}{2}(t-s)}\,\mathrm{e}^{\frac{\lambda^2}{2}(t-s) - \lambda\sigma_S\rho_{Sf}(t-s) + \frac{\sigma_S^2}{2}(t-s)} \\ &= \mathrm{e}^{\frac{\lambda^2}{2}(t-s)}\,\mathrm{e}^{-\lambda q(t-s) - \lambda\sigma_S\rho_{Sf}(t-s)} \\ &= \mathrm{e}^{\frac{\lambda^2}{2}(t-s)} \end{aligned} \tag{A.3}$$

the last equality is ensured if and only if

$$0 = -\lambda q(t-s) - \lambda\sigma_S\rho_{Sf}(t-s) \iff q = -\sigma_S\rho_{Sf}. \tag{A.4}$$

Hence, W^f defined by

$$dW_s^f = d\widetilde{W}_s^f + \rho_{Sf}\sigma_S ds$$

is a $\mathbb{P}$-Brownian motion.

B FX Derivatives - European Call

The payoff at maturity t of a European Call on FX rate is given by

$$(S_t - K)_+$$

with K the strike and S_t the FX rate at time t.

Our aim will be to evaluate V_0

$$V_0 = \mathbb{E}\left[\mathrm{e}^{-\int_0^t r_s^d ds} (S_t - K)_+ \right].$$

Proposition B.1. *If we consider a 3-factor model on Foreign Exchange and Zero-coupon as defined in (1.1), V_0 is given by*[3]

$$V_0 = S_0 P^f(0,t) \mathcal{N}\left(\frac{\log\left(\frac{S_0 P^f(0,t)}{K P^d(0,t)}\right) + \mu(0,t)}{\sigma(0,t)} \right) - K P^d(0,t) \mathcal{N}\left(\frac{\log\left(\frac{S_0 P^f(0,t)}{K P^d(0,t)}\right) - \mu(0,t)}{\sigma(0,t)} \right)$$

with

$$\begin{aligned} \mu(0,t) = & \int_0^t \frac{1}{2} \big(\sigma_S^2(s) + \sigma_f^2(s,t) + \sigma_d^2(s,t) \big) ds \\ & + \int_0^t \big(\rho_{Sf} \sigma_S(s) \sigma_f(s,t) - \rho_{Sd} \sigma_S(s) \sigma_d(s,t) ds - \rho_{fd} \sigma_f(s,t) \sigma_d(s,t) \big) ds \end{aligned}$$

and

$$\sigma^2(0,t) = 2\mu(0,t).$$

Proof. In this part, we want to evaluate

$$V_0 = \mathbb{E}\left[\mathrm{e}^{-\int_0^t r_s^d ds} (S_t - K)_+ \right].$$

If we consider a 3-factor model on Foreign Exchange and Zero-coupon as defined in (1.1), we have

$$\begin{aligned} V_0 &= \mathbb{E}\left[\mathrm{e}^{-\int_0^t r_s^d ds} (S_t - K)_+ \right] \\ &= \mathbb{E}\left[\left(\mathrm{e}^{-\int_0^t r_s^d ds} S_t - \mathrm{e}^{-\int_0^t r_s^d ds} K \right)_+ \right] \\ &= \mathbb{E}\left[\left(\mathrm{e}^{-\int_0^t r_s^d ds} S_t - \mathrm{e}^{-\int_0^t r_s^d ds} K \right) \mathbb{1}_{\{S_t \geqslant K\}} \right] \\ &= \mathbb{E}\left[\mathrm{e}^{-\int_0^t r_s^d ds} S_t \mathbb{1}_{\{S_t \geqslant K\}} \right] - K\,\mathbb{E}\left[\mathrm{e}^{-\int_0^t r_s^d ds} \mathbb{1}_{\{S_t \geqslant K\}} \right]. \end{aligned}$$

We focus on the first term

$$K\,\mathbb{E}\left[\mathrm{e}^{-\int_0^t r_s^d ds} \mathbb{1}_{\{S_t \geqslant K\}} \right]. \tag{B.1}$$

We do the following change of numéraire:

$$\frac{d\widetilde{\mathbb{Q}}}{d\mathbb{P}} = \frac{\widetilde{Z}_t}{\widetilde{Z}_0}$$

with

$$\begin{cases} \widetilde{Z}_t = \exp\left(\widetilde{Y}_t - \frac{1}{2} < \widetilde{Y}, \widetilde{Y} >_t \right), \\ \widetilde{Z}_0 = 1 \end{cases}$$

where $\widetilde{Y}_t = \int_0^t \sigma_d(s,t) dW_s^d$ and $< \widetilde{Y}, \widetilde{Y} >_t = \int_0^t \sigma_d^2(s,t) ds$.

Hence, we can define the following Brownian Motions $\widetilde{W}^d$, $\widetilde{W}^f$, $\widetilde{W}^S$ under $\widetilde{\mathbb{Q}}$:

$$\begin{aligned} d\widetilde{W}_s^d &= dW_s^d - d < Y, W^d >_s &&= dW_s^d - \sigma_d(s,t) ds, \\ d\widetilde{W}_s^f &= dW_s^f - d < Y, W^f >_s &&= dW_s^f - \rho_{fd} \sigma_d(s,t) ds, \\ d\widetilde{W}_s^S &= dW_s^S - d < Y, W^S >_s &&= dW_s^S - \rho_{Sd} \sigma_d(s,t) ds \end{aligned}$$

[3] We ignore the settlements details in the present paper in order to alleviate the notations but the formula can easily be extended to take them into account.

and S_t becomes

$$
\begin{aligned}
S_t &= S_0 \exp\left(\int_0^t \left(r_s^d - r_s^f - \frac{\sigma_S^2(s)}{2}\right)ds + \int_0^t \sigma_S(s)dW_s^S\right) \\
&= \frac{S_0 P^f(0,t)}{P^d(0,t)} \exp\left(\int_0^t -\frac{1}{2}\big(\sigma_S^2(s) + \sigma_f^2(s,t) - \sigma_d^2(s,t)\big) - \rho_{Sf}\sigma_S(s)\sigma_f(s,t)\ ds\right) \\
&\quad \times \exp\left(\int_0^t \sigma_S(s)dW_s^S + \int_0^t \sigma_f(s,t)dW_s^f - \int_0^t \sigma_d(s,t)dW_s^d\right) \\
&= \frac{S_0 P^f(0,t)}{P^d(0,t)} \exp\left(-\int_0^t \frac{1}{2}\big(\sigma_S^2(s) + \sigma_f^2(s,t) + \sigma_d^2(s,t)\big)\ ds\right) \\
&\quad \times \exp\left(-\int_0^t \big(\rho_{Sf}\sigma_S(s)\sigma_f(s,t) - \rho_{Sd}\sigma_S(s)\sigma_d(s,t) - \rho_{fd}\sigma_f(s,t)\sigma_d(s,t)\big)ds\right) \\
&\qquad \times \exp\left(\int_0^t \sigma_S(s)d\widetilde{W}_s^S + \int_0^t \sigma_f(s,t)d\widetilde{W}_s^f - \int_0^t \sigma_d(s,t)d\widetilde{W}_s^d\right) \\
&= \frac{S_0 P^f(0,t)}{P^d(0,t)} \exp\left(-\mu(0,t) + \int_0^t \sigma_S(s)d\widetilde{W}_s^S + \int_0^t \sigma_f(s,t)d\widetilde{W}_s^f - \int_0^t \sigma_d(s,t)d\widetilde{W}_s^d\right).
\end{aligned}
$$

Hence, as $\exp\left(-\int_0^t r_s^d ds\right) = P^d(0,t) \times \widetilde{Z}_t$, (B.1) becomes

$$
\begin{aligned}
K\,\mathbb{E}\left[\mathrm{e}^{-\int_0^t r_s^d ds}\,\mathbb{1}_{\{S_t \geqslant K\}}\right] &= KP^d(0,t)\,\mathbb{E}^{\widetilde{\mathbb{Q}}}\left[\mathbb{1}_{\{S_t \geqslant K\}}\right] \\
&= KP^d(0,t)\widetilde{\mathbb{Q}}(S_t \geqslant K) \\
&= KP^d(0,t)\widetilde{\mathbb{Q}}\left(Z \geqslant \frac{\log\left(\frac{KP^d(0,t)}{S_0P^f(0,t)}\right) + \mu(0,t)}{\sigma(0,t)}\right) \\
&= KP^d(0,t)\widetilde{\mathbb{Q}}\left(Z \leqslant \frac{\log\left(\frac{S_0P^f(0,t)}{KP^d(0,t)}\right) - \mu(0,t)}{\sigma(0,t)}\right) \\
&= KP^d(0,t)\,\mathcal{N}\left(\frac{\log\left(\frac{S_0P^f(0,t)}{KP^d(0,t)}\right) - \mu(0,t)}{\sigma(0,t)}\right)
\end{aligned}
$$

where $Z \sim \mathcal{N}(0,1)$ with

$$
\begin{aligned}
\mu(0,t) &= \int_0^t \frac{1}{2}\big(\sigma_S^2(s) + \sigma_f^2(s,t) + \sigma_d^2(s,t)\big)ds \\
&\quad + \int_0^t \big(\rho_{Sf}\sigma_S(s)\sigma_f(s,t) - \rho_{Sd}\sigma_S(s)\sigma_d(s,t)ds - \rho_{fd}\sigma_f(s,t)\sigma_d(s,t)\big)ds, \\
\sigma^2(0,t) &= \mathbb{V}\mathrm{ar}\left(\int_0^t \sigma_S(s)d\widetilde{W}_s^S + \int_0^t \sigma_f(s,t)d\widetilde{W}_s^f - \int_0^t \sigma_d(s,t)d\widetilde{W}_s^d\right) \\
&= \mathbb{V}\mathrm{ar}\left(\int_0^t \sigma_S(s)d\widetilde{W}_s^S\right) + \mathbb{V}\mathrm{ar}\left(\int_0^t \sigma_f(s,t)d\widetilde{W}_s^f\right) + \mathbb{V}\mathrm{ar}\left(\int_0^t \sigma_d(s,t)d\widetilde{W}_s^d\right) \\
&\quad + 2\,\mathbb{C}\mathrm{ov}\left(\int_0^t \sigma_S(s)d\widetilde{W}_s^S, \int_0^t \sigma_f(s,t)d\widetilde{W}_s^f\right) - 2\,\mathbb{C}\mathrm{ov}\left(\int_0^t \sigma_S(s)d\widetilde{W}_s^S, \int_0^t \sigma_d(s,t)d\widetilde{W}_s^d\right) \\
&\qquad - 2\,\mathbb{C}\mathrm{ov}\left(\int_0^t \sigma_f(s,t)d\widetilde{W}_s^f, \int_0^t \sigma_d(s,t)d\widetilde{W}_s^d\right) \\
&= \int_0^t \big(\sigma_S^2(s) + \sigma_f^2(s,t) + \sigma_d^2(s,t)\big)ds \\
&\quad + 2\int_0^t \big(\rho_{Sf}\sigma_S(s)\sigma_f(s,t) - \rho_{Sd}\sigma_S(s)\sigma_d(s,t) - \rho_{fd}\sigma_f(s,t)\sigma_d(s,t)\big)ds.
\end{aligned}
$$

Now, we deal with the term

$$\mathbb{E}\left[\mathrm{e}^{-\int_0^t r_s^d ds}\, S_t\, \mathbb{1}_{\{S_t \geqslant K\}}\right] = P^d(0,t)\, \mathbb{E}^{\tilde{\mathbb{Q}}}\left[S_t\, \mathbb{1}_{\{S_t \geqslant K\}}\right] \tag{B.2}$$

using directly the formula of the first partial moment of a log-normal random variable. Let $X \sim$ Log-$\mathcal{N}(\mu, \sigma^2)$, then

$$\mathbb{E}\left[X\, \mathbb{1}_{\{X \geqslant x\}}\right] = \mathrm{e}^{\mu + \frac{\sigma^2}{2}}\, \mathcal{N}\left(\frac{\mu + \sigma^2 - \log(x)}{\sigma}\right).$$

Finally, as $S_t = \frac{S_0 P^f(0,t)}{P^d(0,t)} X$ with $X \overset{\tilde{\mathbb{Q}}}{\sim}$ Log-$\mathcal{N}(-\mu(0,t), \sigma^2(0,t))$, we get

$$\begin{aligned}
(\text{B.2}) &= S_0 P^f(0,t)\, \mathbb{E}^{\tilde{\mathbb{Q}}}\left[X\, \mathbb{1}_{\left\{X \geqslant \frac{K P^d(0,t)}{S_0 P^f(0,t)}\right\}}\right] \\
&= S_0 P^f(0,t)\, \mathrm{e}^{-\mu(0,t) + \frac{\sigma^2(0,t)}{2}}\, \mathcal{N}\left(\frac{-\mu(0,t) + \sigma^2(0,t) - \log\left(\frac{K P^d(0,t)}{S_0 P^f(0,t)}\right)}{\sigma(0,t)}\right) \\
&= S_0 P^f(0,t)\, \mathcal{N}\left(\frac{\log\left(\frac{S_0 P^f(0,t)}{K P^d(0,t)}\right) + \mu(0,t)}{\sigma(0,t)}\right)
\end{aligned}$$

noticing that $\mu(0,t) = \frac{\sigma^2(0,t)}{2}$.

Finally, we get

$$\begin{aligned}
V_0 &= \mathbb{E}\left[\mathrm{e}^{-\int_0^t r_s^d ds}(S_t - K)_+\right] \\
&= \mathbb{E}\left[\mathrm{e}^{-\int_0^t r_s^d ds}\, S_t\, \mathbb{1}_{\{S_t \geqslant K\}}\right] - K\, \mathbb{E}\left[\mathrm{e}^{-\int_0^t r_s^d ds}\, \mathbb{1}_{\{S_t \geqslant K\}}\right] \\
&= S_0 P^f(0,t)\, \mathcal{N}\left(\frac{\log\left(\frac{S_0 P^f(0,t)}{K P^d(0,t)}\right) + \mu(0,t)}{\sigma(0,t)}\right) - K P^d(0,t)\, \mathcal{N}\left(\frac{\log\left(\frac{S_0 P^f(0,t)}{K P^d(0,t)}\right) - \mu(0,t)}{\sigma(0,t)}\right).
\end{aligned}$$

Special case of constant volatility: $\sigma_S(s) = \sigma_S$, $\sigma_d(s,t) = \sigma_d \times (t-s)$ $\sigma_f(s,t) = \sigma_f \times (t-s)$

$$\begin{aligned}
\mu(0,t) &= \int_0^t \frac{1}{2}\big(\sigma_S^2(s) + \sigma_f^2(s,t) + \sigma_d^2(s,t)\big) ds \\
&\quad + \int_0^t \big(\rho_{Sf}\sigma_S(s)\sigma_f(s,t) - \rho_{Sd}\sigma_S(s)\sigma_d(s,t) - \rho_{fd}\sigma_f(s,t)\sigma_d(s,t)\big) ds \\
&= \int_0^t \frac{1}{2}\big(\sigma_S^2 + \sigma_f^2(t-s)^2 + \sigma_d^2(t-s)^2\big) ds \\
&\quad + \int_0^t \rho_{Sf}\sigma_S\sigma_f(t-s) - \rho_{Sd}\sigma_S\sigma_d(t-s) - \rho_{fd}\sigma_f\sigma_d(t-s)^2 ds \\
&= \frac{1}{2}\left(\sigma_S^2 t + \sigma_f^2\frac{t^3}{3} + \sigma_d^2\frac{t^3}{3}\right) + \rho_{Sf}\sigma_S\sigma_f\frac{t^2}{2} - \rho_{Sd}\sigma_S\sigma_d\frac{t^2}{2} - \rho_{fd}\sigma_f\sigma_d\frac{t^3}{3}, \\
\sigma^2(0,t) &= \int_0^t \big(\sigma_S^2(s) + \sigma_f^2(s,t) + \sigma_d^2(s,t)\big) ds \\
&\quad + 2\int_0^t \big(\rho_{Sf}\sigma_S(s)\sigma_f(s,t) - \rho_{Sd}\sigma_S(s)\sigma_d(s,t) - \rho_{fd}\sigma_f(s,t)\sigma_d(s,t)\big) ds \\
&= 2\mu(0,t).
\end{aligned}$$

$\square$