

# Quantization-based Bermudan option pricing in the FX world

Jean-Michel Fayolle, Vincent Lemaire, Thibaut Montes, Gilles Pagès

▶ **To cite this version:**

Jean-Michel Fayolle, Vincent Lemaire, Thibaut Montes, Gilles Pagès. Quantization-based Bermudan option pricing in the FX world. 2020. hal-02361667v2

**HAL Id: hal-02361667**
**https://hal.science/hal-02361667v2**

Preprint submitted on 1 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Quantization-based Bermudan option pricing in the $FX$ world

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

May 1, 2020

**Abstract**

This paper proposes two numerical solution based on Product Optimal Quantization for the pricing of Foreign Exchange (FX) linked long term Bermudan options e.g. 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

$$\frac{dP(t,T)}{P(t,T)} = r_t dt + \sum_i \sigma_i\big(t,T,P(t,T)\big)dW_t^i \tag{0.1}$$

where $r_t$ is the instantaneous rate at time $t$ (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

$$\frac{dP(t,T)}{P(t,T)} = r_t dt + \sigma(T-t)dW_t \tag{0.2}$$

[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, $\sigma$ 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 (0.2) 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 one. The diffusion is that of a standard Black-Scholes model with the following equation

$$\frac{dS_t}{S_t} = (r_t^d - r_t^f)dt + \sigma_S dW_t^S \tag{0.3}$$

where $r_t^d$ is the instantaneous rate of the domestic currency at time $t$, $r_t^f$ is the instantaneous rate of the foreign currency at time $t$, $\sigma_S$ is the volatility of the $FX$ rate and $W^S$ is a standard Brownian motion under the 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 $\Gamma_N$, a finite subset of $\mathbb{R}^d$ defined by $\Gamma_N := \{z_1^N, \dots, z_N^N\} \subset \mathbb{R}^d$, 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 $\Gamma_N$. 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 $Z^\ell$ independently by building a one-dimensional optimal quantization $\widehat{Z}^\ell$ of size $N^\ell$, with quantizer $\Gamma_\ell^{N_\ell} = \{z_{i_\ell}^\ell, i_\ell \in \{1, \cdots, N_\ell\}\}$ 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

$$\Gamma^N = \{(z_{i_1}^1, \cdots, z_{i_\ell}^\ell, \cdots, z_{i_d}^d), \quad i_\ell \in \{1, \cdots, N_\ell\}, \quad \ell \in \{1, \cdots, d\}\}. \tag{0.4}$$

Then, in the 90s, [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

$$\lim_{N\to+\infty} N^{\alpha} \big| \mathbb{E}\big[f(Z)\big] - \mathbb{E}\big[f(\widehat{Z}^N)\big]\big| \leqslant C_{f,X} < +\infty. \tag{0.5}$$

For more results on the rate of convergence, the value of $\alpha$, 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 deal 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

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

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 $\sigma_d$ 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 $\sigma_f$ is the volatility for the foreign zero coupon curve. 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 $\rho_{df}$ defined as limit of the quadratic variation $\langle W^d, \widetilde{W}^f \rangle_t = \rho_{df} t$.

**Remarks 1.1.** 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

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

with $W_t^S$ a $\mathbb{P}$-Brownian Motion under the domestic risk-neutral probability such that their exist $\rho_{Sd}$ and $\rho_{Sf}$ defined as limit of the quadratic variation $\langle W^S, W^d \rangle_t = \rho_{Sd} t$ 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. Using Itô's formula, we can explicitly express the processes

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

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

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

where

$$\varphi_d(t) = P^d(0,t)\exp\left(-\sigma_d^2\int_0^t \frac{(t-s)^2}{2}ds\right) \tag{1.2}$$

and

$$\varphi_f(t) = P^f(0,t)\exp\left(-\int_0^t \left(\rho_{Sf}\sigma_S\sigma_f(t-s) + \frac{\sigma_f^2(t-s)^2}{2}\right)ds\right). \tag{1.3}$$

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 $(\Omega, \mathcal{A}, \mathbb{P})$ our domestic risk neutral probability space. We want to evaluate the price of a Bermudan option on the $FX$ rate $S_t$ defined by

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

with

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

where the owner of the financial product can exercise its option at predetermined dates $t_0, t_1, \cdots, t_n$ with payoff $\psi_{t_k}$ at date $t_k$, where $t_0 = 0$.

At a given time $t$, the observables in the market are the foreign exchange rate $S_t$ and the zero-coupon curves $\left(P^d(t,T)\right)_{T\geqslant t}$ and $\left(P^f(t,T)\right)_{T\geqslant t}$, hence the natural filtration to consider is

$$\mathcal{F}_t = \sigma\big(S_s, P^d(s,T), P^f(s,T), s\leqslant t\big) = \sigma\big(W_s^S, W_s^d, W_s^f, s\leqslant t\big). \tag{2.1}$$

Let $\tau : \Omega \to \{t_0, t_1, \dots, t_n\}$ a stopping time for the filtration $(\mathcal{F}_{t_k})_{k\geqslant 0}$ and $\mathcal{T}$ the set of all stopping times for the filtration $(\mathcal{F}_{t_k})_{k\geqslant 0}$. In this paper, we consider problems where the horizon is finite then we define $\mathcal{T}_k^n$, the set of all stopping times taking finite values

$$\mathcal{T}_k^n = \big\{\tau \in \mathcal{T}, \mathbb{P}(t_k \leqslant \tau \leqslant t_n) = 1\big\}. \tag{2.2}$$

Hence, the price at time $t_k$ of the Bermudan option is given by

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

and $V_k$ is called the *Snell envelope* of the obstacle process $\left( \mathrm{e}^{-\int_0^{t_k} r_s^d ds}\, \psi_{t_k}(S_{t_k}) \right)_{k=0:n}$ such that

$$\mathbb{E}\left[\psi_{t_k}(S_{t_k})^2\right] < +\infty, \quad \forall k = 0, \dots, n. \tag{2.3}$$

**Remark 2.1.** The financial products we consider in the applications are PRDC. Their payoffs (see Figure 1) have the following expression

$$\psi_{t_k}(x) = \min\left( \max\left( \frac{C_f(t_k)}{S_0} x - C_d(t_k), \mathrm{Floor}(t_k) \right), \mathrm{Cap}(t_k) \right) \tag{2.4}$$

where $\mathrm{Floor}(t_k)$ and $\mathrm{Cap}(t_k)$ are the floor and cap values chosen at the creation of the product, as well as $C_f(t_k)$ and $C_d(t_k)$ that are the coupons value we wish to compare to the foreign and the domestic currency, respectively.

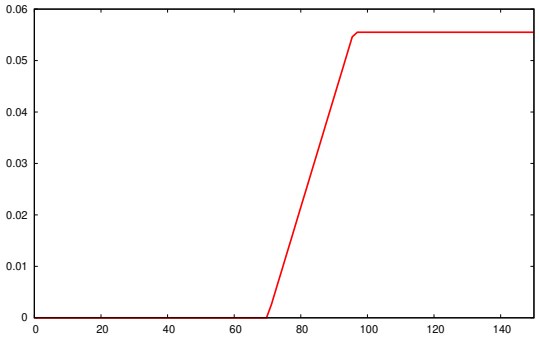



Figure 1: *Example of a PRDC payoff* $\psi_{t_k}(S_{t_k}) = \min\left( \left( 0.189 \frac{S_{t_k}}{88.17} - 0.15 \right)_+, 0.0555 \right)$ *at time* $t_k$.

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

## 2.2 Backward Dynamic Programming Principle

$V_k$ can also be defined recursively by

$$\begin{cases} V_n = \mathrm{e}^{-\int_0^{t_n} r_s^d ds}\, \psi_n(S_{t_n}), \\ V_k = \max\left( \mathrm{e}^{-\int_0^{t_k} r_s^d ds}\, \psi_k(S_{t_k}), \mathbb{E}[V_{k+1} \mid \mathcal{F}_{t_k}] \right), \qquad 0 \leqslant k \leqslant n-1 \end{cases} \tag{2.5}$$

and this representation is called the *Backward Dynamic Programming Principle* (BDPP).

First, noticing that the obstacle process $\mathrm{e}^{-\int_0^t r_s^d ds}\, \psi_t(S_t)$ can be rewritten 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 $h$ is given by

$$h_t(x,y) = \varphi_d(t)\,\mathrm{e}^{-y}\,\psi_t\left(S_0\frac{\varphi_f(t)}{\varphi_d(t)}\,\mathrm{e}^{-\sigma_S^2 t/2+x+y}\right) \tag{2.6}$$

and $(X,Y)$ is defined by

$$(X_t,Y_t) = \left(\sigma_S W_t^S + \sigma_f\int_0^t (t-s)dW_s^f, -\sigma_d\int_0^t (t-s)dW_s^d\right). \tag{2.7}$$

Now, 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 $h_k = h_{t_k}$.

Using this new form, the Snell envelope becomes

$$V_k = \sup_{\tau\in\mathcal{T}_k^n} \mathbb{E}\left[h_\tau(X_\tau,Y_\tau)\mid\mathcal{F}_{t_k}\right]$$

and the *Backward Dynamic Programming Principle* (2.5) rewrites

$$\begin{cases} V_n = h_n(X_n,Y_n), \\ V_k = \max\left(h_k(X_k,Y_k), \mathbb{E}\left[V_{k+1}\mid\mathcal{F}_{t_k}\right]\right), \qquad 0\leqslant k\leqslant n-1. \end{cases} \tag{2.8}$$

Second, in order to solve the problem theoretically by dynamic programming it is required to associate a $\mathcal{F}_t$-Markov process to this problem and in our case, the simplest of them (i.e. of minimal dimension) is $(X_t, W_t^f, Y_t, W_t^d)$ which is $\mathcal{F}_t$-adapted and a Markov process because

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

where $\delta = \frac{T}{n}$ and can be written as

$$\begin{cases} X_{k+1} = X_k + \sigma_f\delta W_k^f + G_{k+1}^1 \\ W_{k+1}^f = W_k^f + G_{k+1}^2 \\ Y_{k+1} = Y_k - \sigma_d\delta W_k^d + G_{k+1}^3 \\ W_{k+1}^d = W_k^d + G_{k+1}^4, \end{cases} \tag{2.9}$$

where the increments are normally distributed

$$\begin{pmatrix} G_{k+1}^1 \\ G_{k+1}^2 \\ G_{k+1}^3 \\ G_{k+1}^4 \end{pmatrix} \sim \mathcal{N}\left(\mu_{k+1}, \Sigma_{k+1}\right) \tag{2.10}$$

with

$$\mu_{k+1} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad \text{and} \qquad \Sigma_{k+1} = \left(\mathbb{C}\mathrm{ov}\left(G_{k+1}^i, G_{k+1}^j\right)\right)_{i,j=1:4}. \tag{2.11}$$

One notices that $\big((G_k^1, G_k^2, G_k^3, G_k^4)\big)_{k=1\dots,n}$ are i.i.d. Based on Equation (2.9), we deduce the Markov process transition of $(X_k, W_k^f, Y_k, W_k^d)$, for any integrable function $f : \mathbb{R}^4 \to \mathbb{R}$, given by

$$Pf(x,u,y,v) = \mathbb{E}\big[f(x + \sigma_f \delta u + G_{k+1}^1, u + G_{k+1}^2, y - \sigma_d \delta v + G_{k+1}^3, v + G_{k+1}^4)\big]. \tag{2.12}$$

**Remark 2.2.** Using the Markov process $(X, W^f, Y, W^d)$ newly defined, we rewrite the filtration $\mathcal{F}_t$ as

$$\mathcal{F}_t = \sigma\big(W_s^S, W_s^d, W_s^f, s \leqslant t\big) = \sigma\big(X_s, W_s^f, Y_s, W_s^d, s \leqslant t\big). \tag{2.13}$$

Then, using the new expression for the filtration and the Markov property of $(X_k, W_k^f, Y_k, W_k^d)$, the BDPP (2.8) reads as follows,

$$\begin{cases} V_n = h_n(X_n, Y_n), \\ V_k = \max\Big(h_k(X_k, Y_k), \mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big]\Big), \qquad 0 \leqslant k \leqslant n-1. \end{cases} \tag{2.14}$$

Moreover, by backward induction we get $V_k = v_k(X_k, W_k^f, Y_k, W_k^d)$ where

$$\begin{cases} v_n(X_n, W_n^f, Y_n, W_n^d) = h_n(X_n, Y_n), \\ v_k(X_k, W_k^f, Y_k, W_k^d) = \max\Big(h_k(X_k, Y_k), Pv_{k+1}(X_k, W_k^f, Y_k, W_k^d)\Big), \qquad 0 \leqslant k \leqslant n-1. \end{cases} \tag{2.15}$$

**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 2.3.** *If $\psi_{t_k}$ is are Lipschitz continuous with Lipschitz coefficient $[\psi_{t_k}]_{Lip}$ with compactly supported (right) derivative (such as the payoff defined in (2.4)) then $h_k(x,y)$ given by (2.6) is locally Lipschitz continuous, for every $x, x', y, y' \in \mathbb{R}$*

$$|h_k(x,y) - h_k(x',y')| \leqslant \mathrm{e}^{|y| \vee |y'|} \left([\bar{\psi}_k]_{Lip}|x - x'| + (\varphi_d(t_k)\|\psi_{t_k}\|_\infty + [\bar{\psi}_k]_{Lip})|y - y'|\right) \tag{2.16}$$

*with $[\bar{\psi}_k]_{Lip} = [\psi_{t_k}]_{Lip} S_0 \varphi_f(t_k)\, \mathrm{e}^{-\sigma_S^2 t_k/2}\, \|\psi'_{t_k}\|_\infty\, \mathrm{e}^c$ with $\psi'_{t_k}$ the right derivative of $\psi_{t_k}$.*

*Proof.* Let $g_k$ be defined by

$$g_k(x,y) = \psi_{t_k}\left(S_0 \frac{\varphi_f(t_k)}{\varphi_d(t_k)}\, \mathrm{e}^{-\sigma_S^2 t_k/2 + x + y}\right). \tag{2.17}$$

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

$$|(\psi_{t_k}(\mathrm{e}^x))'| = |\,\mathrm{e}^x\, \psi'_{t_k}(\mathrm{e}^x)| \leqslant \|\psi'_{t_k}\|_\infty \sup_{x \in \operatorname{supp} \psi'_{t_k}} \mathrm{e}^x \leqslant \|\psi'_{t_k}\|_\infty\, \mathrm{e}^c\,. \tag{2.18}$$

Hence

$$|g_k(x,y) - g_k(x',y')| \leqslant \frac{[\bar{\psi}_k]_{Lip}}{\varphi_d(t_k)}\big(|x - x'| + |y - y'|\big) \tag{2.19}$$

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

$$\begin{aligned} |h_k(x,y) - h_k(x',y')| &= \varphi_d(t_k)\big|\,\mathrm{e}^{-y}\, g_k(x,y) - \mathrm{e}^{-y'}\, g_k(x',y')\big| \\ &\leqslant \varphi_d(t_k)\Big(\big|\,\mathrm{e}^{-y}\, g_k(x,y) - \mathrm{e}^{-y'}\, g_k(x,y)\big| + \big|\,\mathrm{e}^{-u'}\, g_k(x,y) - \mathrm{e}^{-y'}\, g_k(x',y')\big|\Big) \\ &\leqslant \varphi_d(t_k)\Big(\big|\,\mathrm{e}^{-y} - \mathrm{e}^{-y'}\,\big| \cdot \|\psi_{t_k}\|_\infty + \mathrm{e}^{-y'}\,\big|g_k(x,y) - g_k(x',y')\big|\Big) \\ &\leqslant \mathrm{e}^{|y| \vee |y'|}\left([\bar{\psi}_k]_{Lip}|x - x'| + (\varphi_d(t_k)\|\psi_{t_k}\|_\infty + [\bar{\psi}_k]_{Lip})|y - y'|\right). \end{aligned} \tag{2.20}$$

□

The next Lemma shows that the transition of the Markov process propagates the local Lipschitz continuity of a function $f$. This result will be helpful to estimate the error induced by the numerical approximation (2.15).

**Lemma 2.4.** *Let* $Pf(x,u,y,v) = \mathbb{E}\big[f(x+\sigma_f\delta u + G^1, u+G^2, y-\sigma_d\delta v + G^3, v+G^4)\big]$ *be a Markov kernel. If the function* $f$ *satisfies the following local Lipschitz property,*

$$|f(x,u,y,v) - f(x',u',y',v')| \leqslant \big(A|x-x'| + B|u-u'| + C|y-y'| + D|v-v'|\big) \times \mathrm{e}^{|y|\vee|y'|+b|v|\vee|v'|} \tag{2.21}$$

*then*

$$|Pf(x,u,y,v) - Pf(x',u',y',v')| \leqslant \big(\widetilde{A}|x-x'| + \widetilde{B}|u-u'| + \widetilde{C}|y-y'| + \widetilde{D}|v-v'|\big) \times \mathrm{e}^{|y|\vee|y'|+\widetilde{b}|v|\vee|v'|}. \tag{2.22}$$

*Proof.* It follows from Jensen's inequality and our assumption on $f$

$$\begin{aligned}
&|Pf(x,u,y,v) - Pf(x',u',y',v')| \\
&\quad\leqslant \mathbb{E}\Big[\big|f(x+\sigma_f\delta u + G^1, u+G^2, y-\sigma_d\delta v+G^3, v+G^4) \\
&\qquad\qquad - f(x'+\sigma_f\delta u' + G^1, u'+G^2, y'-\sigma_d\delta v'+G^3, v'+G^4)\big|\Big] \\
&\quad\leqslant \big(A|x-x'| + (B+A\sigma_f\delta)|u-u'| + C|y-y'| + (D+C\sigma_d\delta)|v-v'|\big) \\
&\qquad\qquad \times \mathrm{e}^{|y|\vee|y'|+(b+\sigma_d\delta)|v|\vee|v'|}\,\mathbb{E}\big[\,\mathrm{e}^{|G^3|+b|G^4|}\,\big] \\
&\quad\leqslant \big(\widetilde{A}|x-x'| + \widetilde{B}|u-u'| + \widetilde{C}|y-y'| + \widetilde{D}|v-v'|\big) \\
&\qquad\qquad \times \mathrm{e}^{|y|\vee|y'|+\widetilde{b}|v|\vee|v'|}
\end{aligned} \tag{2.23}$$

where

$$\widetilde{A} = A\,\mathbb{E}[\kappa], \qquad \widetilde{B} = (B + A\sigma_f\delta)\,\mathbb{E}[\kappa] \tag{2.24}$$

and

$$\widetilde{C} = C\,\mathbb{E}[\kappa], \qquad \widetilde{D} = (D + C\sigma_d\delta)\,\mathbb{E}[\kappa], \qquad \widetilde{b} = b + \sigma_d\delta \tag{2.25}$$

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

**Value function regularity.** If the functions $(\psi_{t_k})_{k=0:n}$ are defined as in Equation (2.4) then $v_n(x,u,y,v)$ preserves a local Lipschitz property. Hence, for every $x,x',u,u',y,y',v,v' \in \mathbb{R}$,

$$|v_n(x,u,y,v) - v_n(x',u',y',v')| \leqslant \big(A_n|x-x'| + B_n|u-u'| + C_n|y-y'| + D_n|v-v'|\big) \times \mathrm{e}^{|y|\vee|y'|+b_n|v|\vee|v'|} \tag{2.26}$$

where

$$A_n = [\bar{\psi}_n]_{Lip}, \qquad B_n = 0, \qquad C_n = \varphi_d(t_n)\|\psi_n\|_\infty + [\bar{\psi}_n]_{Lip}, \qquad D_n = 0, \qquad b_n = 0 \tag{2.27}$$

with $[\bar{\psi}_n]_{Lip} = [\psi_{t_n}]_{Lip} S_0 \varphi_f(t_n) \exp(-\sigma_S^2 t_n/2)\|\psi'_{t_n}\|_\infty\,\mathrm{e}^c$. Using now Lemma 2.4 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

$$
\begin{aligned}
&|v_k(x,u,y,v) - v_k(x',u',y',v')| \\
&\qquad \leqslant \max\big(|h_k(x,y) - h_k(x',y')|, |Pv_{k+1}(x,u,y,v) - Pv_{k+1}(x',u',y',v')|\big) \\
&\qquad \leqslant \max\Bigg( \mathrm{e}^{|y|\vee|y'|}\left([\bar{\psi}_k]_{Lip}|x-x'| + \left(\varphi_d(t_k)\|\psi_{t_k}\|_\infty + [\bar{\psi}_k]_{Lip}\right)|y-y'|\right) \\
&\qquad\qquad , \left(\widetilde{A}_k|x-x'| + \widetilde{B}_k|u-u'| + \widetilde{C}_k|y-y'| + \widetilde{D}_k|v-v'|\right) \\
&\qquad\qquad \times \mathrm{e}^{|y|\vee|y'|+\widetilde{b}_k|v|\vee|v'|} \Bigg) \\
&\qquad \leqslant \left(A_k|x-x'| + B_k|u-u'| + C_k|y-y'| + D_k|v-v'|\right) \\
&\qquad\qquad \times \mathrm{e}^{|y|\vee|y'|+b_k|v|\vee|v'|}
\end{aligned}
\tag{2.28}
$$

where

$$
A_k = [\bar{\psi}_k]_{Lip} \vee \left(A_{k+1}\,\mathbb{E}[\kappa_{k+1}]\right), \qquad B_k = \left(B_{k+1} + A_{k+1}\sigma_f\delta\right)\mathbb{E}[\kappa_{k+1}], \qquad b_k = b_{k+1} + \sigma_d\delta \tag{2.29}
$$

and

$$
C_k = \left(\varphi_d(t_k)\|\psi_{t_k}\|_\infty + [\bar{\psi}_k]_{Lip}\right) \vee \left(C_{k+1}\,\mathbb{E}[\kappa_{k+1}]\right), \qquad D_k = \left(D_{k+1} + C_{k+1}\sigma_d\delta\right)\mathbb{E}[\kappa_{k+1}] \tag{2.30}
$$

with $\kappa_{k+1} = \exp(|G^3_{k+1}| + b_{k+1}|G^4_{k+1}|)$. Or equivalently

$$
A_k = \max_{l\geqslant k}\left([\bar{\psi}_l]_{Lip}\prod_{j=k+1}^{l}\mathbb{E}[\kappa_j]\right), \qquad B_k = \sigma_f\frac{T}{n}\sum_{l=k+1}^{n}\max_{l\leqslant i\leqslant n}\left([\bar{\psi}_i]_{Lip}\prod_{j=k+1}^{i}\mathbb{E}[\kappa_j]\right) \tag{2.31}
$$

and

$$
\begin{aligned}
C_k &= \max_{l\geqslant k}\left(\left(\varphi_d(t_l)\|\psi_l\|_\infty + [\bar{\psi}_l]_{Lip}\right)\prod_{j=k+1}^{l}\mathbb{E}[\kappa_j]\right), \\
D_k &= \sigma_d\frac{T}{n}\sum_{l=k+1}^{n}\max_{l\leqslant i\leqslant n}\left(\left(\varphi_d(t_i)\|\psi_i\|_\infty + [\bar{\psi}_i]_{Lip}\right)\prod_{j=k+1}^{i}\mathbb{E}[\kappa_j]\right)
\end{aligned}
\tag{2.32}
$$

with

$$
b_k = \sigma_d T\Big(1 - \frac{k-1}{n}\Big). \tag{2.33}
$$

# 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 is an optimal quantizer and what we mean by a product quantization tree. Second, we present a first numerical solution, based on quantization of the Markovian tuple $(X, W^f, Y, W^d)$, to solve the numerical problem (2.14) 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 About Optimal Quantization

**Theoretical background (the one-dimensional case).** The aim of Optimal Quantization is to determine $\Gamma_N$, a set with cardinality at most $N$, which minimises the quantization error among all such sets $\Gamma$. We place ourselves in the one-dimensional case. Let $Z$ be an $\mathbb{R}$-valued random variable with distribution $\mathbb{P}_Z$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that $Z \in L^2_{\mathbb{R}}$.

**Definition 3.1.** Let $\Gamma_N = \{z_1, \dots, z_N\} \subset \mathbb{R}$ be a subset of size $N$, called $N$-quantizer. A Borel partition $\big(C_i(\Gamma_N)\big)_{i\in[\![1,N]\!]}$ of $\mathbb{R}$ is a Voronoï partition of $\mathbb{R}$ induced by the $N$-quantizer $\Gamma_N$ if, for every $i = \{1, \cdots, N\}$,

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

The Borel sets $C_i(\Gamma_N)$ are called Voronoï cells of the partition induced by $\Gamma_N$.

One can always consider that the quantizers are ordered: $z_1 < z_2 < \cdots < z_{N-1} < z_N$ and in that case the Voronoï cells are given by

$$C_k\big(\Gamma_N\big) = \big(z_{k-1/2}, z_{k+1/2}\big], \qquad k \in [\![1, N-1]\!], \qquad C_N\big(\Gamma_N\big) = \big(z_{N-1/2}, z_{N+1/2}\big)$$

where $\forall k \in \{2, \cdots, N\}$, $\quad z_{k-1/2} = \frac{z_{k-1}+z_k}{2}$ and $z_{1/2} = \inf\big(\mathrm{supp}(\mathbb{P}_Z)\big)$ and $z_{N+1/2} = \sup\big(\mathrm{supp}(\mathbb{P}_Z)\big)$.

**Definition 3.2.** Let $\Gamma_N = \{z_1, \dots, z_N\}$ be an $N$-quantizer. The nearest neighbour projection $\mathrm{Proj}_{\Gamma_N} : \mathbb{R} \to \{z_1, \dots, z_N\}$ induced by a Voronoï partition $\big(C_i(\Gamma_N)\big)_{i\in\{1,\cdots,N\}}$ is defined by

$$\forall \xi \in \mathbb{R}, \qquad \mathrm{Proj}_{\Gamma_N}(\xi) = \sum_{i=1}^{N} z_i \, \mathbb{1}_{\xi \in C_i(\Gamma_N)}\,.$$

Hence, we can define the quantization of $Z$ as the nearest neighbour projection of $Z$ onto $\Gamma_N$ by composing $\mathrm{Proj}_{\Gamma_N}$ and $X$

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

In order to alleviate notations, we write $\widehat{Z}^N$ from now on in place of $\widehat{Z}^{\Gamma_N}$.

Now that we have defined the quantization of $Z$, we explain where does the term "optimal" comes from in the term optimal quantization. First, we define the quadratic distortion function.

**Definition 3.3.** The $L^2$-mean quantization error induced by the quantizer $\widehat{Z}^N$ is defined as

$$\|Z - \widehat{Z}^N\|_2 = \left(\mathbb{E}\Big[\min_{i\in\{1,\cdots,N\}} |Z - z_i|^2\Big]\right)^{1/2} = \left(\int_{\mathbb{R}} \min_{i\in\{1,\cdots,N\}} |\xi - z_i|^2 \, \mathbb{P}_Z(d\xi)\right)^{1/2}. \tag{3.1}$$

It is convenient to define the quadratic distortion function at level $N$ as the squared mean quadratic quantization error on $(\mathbb{R})^N$:

$$\mathcal{Q}_{2,\mathrm{N}} : z = \big(z_1, \dots, z_N\big) \longmapsto \mathbb{E}\Big[\min_{i\in\{1,\cdots,N\}} |Z - z_i|^2\Big] = \|Z - \widehat{Z}^N\|_2^2.$$

**Remark 3.4.** All these definitions can be extended to the $L^p$ case. For example the $L^p$-mean quantization error induced by a quantizer of size $N$ is

$$\|Z - \widehat{Z}^N\|_p = \left(\mathbb{E}\Big[\min_{i\in\{1,\cdots,N\}} |Z - z_i|^p\Big]\right)^{1/p} = \left(\int_{\mathbb{R}} \min_{i\in\{1,\cdots,N\}} |Z - z_i|^p \, \mathbb{P}_Z(d\xi)\right)^{1/p}. \tag{3.2}$$

The existence of a $N$-tuple $z^{(N)} = (z_1, \ldots, z_N)$ minimizing the quadratic distortion function $\mathcal{Q}_{2,\mathrm{N}}$ at level $N$ has been shown and its associated quantizer $\Gamma_N = \{z_i, i \in \{1, \cdots, N\}\}$ is called an optimal quadratic $N$-quantizer, see e.g. [Pag18] for further details and references. We now turn to the asymptotic behaviour in $N$ of the quadratic mean quantization error. The next Theorem, known as Zador's Theorem, provides the sharp rate of convergence of the $L^p$-mean quantization error.

**Theorem 3.5.** *(Zador's Theorem) Let $p \in (0, +\infty)$.*

*(a)* Sharp rate. *Let $Z \in L^{p+\delta}_{\mathbb{R}}(\mathbb{P})$ for some $\delta > 0$. Let $\mathbb{P}_Z(d\xi) = \varphi(\xi) \cdot \lambda(d\xi) + \nu(d\xi)$, where $\nu \perp \lambda$ i.e. denotes the singular part of $\mathbb{P}_Z$ with respect to the Lebesgue measure $\lambda$ on $\mathbb{R}$. Then,*

$$\lim_{N \to +\infty} N \min_{\Gamma_N \subset \mathbb{R}, |\Gamma_N| \leqslant N} \|Z - \widehat{Z}^N\|_p = \frac{1}{2^p(p+1)} \left[ \int_{\mathbb{R}} \varphi^{\frac{1}{1+p}} d\lambda \right]^{1+\frac{1}{p}}. \tag{3.3}$$

*(b)* Non asymptotic upper-bound. *Let $\delta > 0$. There exists a real constant $C_{1,p,\delta} \in (0, +\infty)$ such that, for every $\mathbb{R}$-valued random variable $Z$,*

$$\forall N \geqslant 1, \qquad \min_{\Gamma_N \subset \mathbb{R}, |\Gamma_N| \leqslant N} \|Z - \widehat{Z}^N\|_p \leqslant C_{1,p,\delta} \sigma_{\delta+p}(Z) N^{-1} \tag{3.4}$$

*where, for $r \in (0, +\infty)$, $\sigma_r(Z) = \min_{a \in \mathbb{R}} \|Z - a\|_r < +\infty$.*

The next result answers to the following question: what can be said about the convergence rate of $\mathbb{E}\left[|Z - \widehat{Z}^N|^{2+\beta}\right]$, knowing that $\widehat{Z}^N$ is a quadratic optimal quantization?

This problem is known as the distortion mismatch problem and has been first addressed by [GLP08] and the results have been extended in Theorem 4.3 of [PS18].

**Theorem 3.6.** *[$L^r$-$L^s$-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 density $\varphi$. Let $(\Gamma_N)_{N \geqslant 1}$ be a sequence of $L^r$-optimal grids. Let $s \in (r, r+1)$. If*

$$Z \in L^{\frac{s}{1+r-s}+\delta}(\Omega, \mathcal{A}, \mathbb{P}) \tag{3.5}$$

*for some $\delta > 0$, then*

$$\limsup_N N \|Z - \widehat{Z}^N\|_s < +\infty. \tag{3.6}$$

**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 $Z^\ell$ independently, build a one-dimensional optimal quantization $\widehat{Z}^\ell$, of size $N^\ell$, with quantizer $\Gamma_\ell^{N^\ell} = \left\{z_{i_\ell}^\ell, i_\ell \in \{1, \cdots, N^\ell\}\right\}$ and then build the product quantizer $\Gamma^N = \prod_{\ell=1}^d \Gamma_l^{N^\ell}$ of size $N = N^1 \times \cdots \times N^d$ defined by

$$\Gamma^N = \left\{(z_{i_1}^1, \cdots, z_{i_\ell}^\ell, \cdots, z_{i_d}^d), \quad i_\ell \in \{1, \cdots, N_\ell\}, \quad \ell \in \{1, \cdots, d\}\right\}. \tag{3.7}$$

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 it exists 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, in place of considering a random variable $Z$, let $(Z_t)_{t\in[0,T]}$ be a stochastic process following a Stochastic Differential Equation (SDE)

$$Z_t = Z_0 + \int_0^t b_s(Z_s)ds + \int_0^t \sigma(s, Z_s)dW_s \tag{3.8}$$

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 $\sigma$ 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 $\pi_{ij}^k = \mathcal{L}(\widehat{Z}_{k+1} \mid \widehat{Z}_k)$. 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 idea in order to discretize (2.14) is to replace the processes by a product quantizer composed with optimal quadratic quantizers. Indeed, at each time $t_k$, we know the law of the processes $X_k$, $W_k^f$, $Y_k$ and $W_k^d$. Then we "force" in some sense the (lost) Markov property by introducing the *Quantized Backward Dynamic Programming Principle* (QBDPP) defined by

$$\begin{cases} \widehat{V}_n = h_n(\widehat{X}_n, \widehat{Y}_n), \\ \widehat{V}_k = \max\left(h_k(\widehat{X}_k, \widehat{Y}_k), \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\right]\right), \qquad 0 \leqslant k \leqslant n-1, \end{cases} \tag{3.9}$$

where for every $k = 0, \ldots, n$, $\widehat{X}_k$, $\widehat{W}_k^f$, $\widehat{Y}_k$ and $\widehat{W}_k^d$ are quadratic optimal quantizers of $X_k$, $W_k^f$, $Y_k$ and $W_k^d$ of size $N_k^X$, $N_k^{W^f}$, $N_k^Y$ and $N_k^{W^d}$ respectively and we denote $N_k = N_k^X \times N_k^{W^f} \times N_k^Y \times N_k^{W^d}$ the size of the grid of the product quantizer.

We are interested by the error induced by the numerical algorithm defined in (3.9) and more precisely its $L^2$-error, with in mind that we "lost" the Markov property in the quantization procedure. Moreover, this can be circumvented as shown below.

**Theorem 3.7.** *Let the Markov transition $Pf(x,u,y,v)$ defined in (2.12) be locally Lipschitz in the sense of Lemma 2.4. Assume that all the payoff functions $(\psi_{t_k})_{k=0:n}$ are Lipschitz continuous with compactly supported (right) derivative. Then the $L^2$-error induced by the quantization*

*approximation* $(\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)$ *is upper-bounded by*

$$\left\|V_k - \widehat{V}_k\right\|_2 \leqslant \left( \sum_{l=k}^{n} C_{X_l} \left\|X_l - \widehat{X}_l\right\|_{2p}^2 + C_{Y_l} \left\|Y_l - \widehat{Y}_l\right\|_{2p}^2 + C_{W_l^d} \left\|W_l^d - \widehat{W}_l^d\right\|_{2p}^2 + C_{W_l^f} \left\|W_l^f - \widehat{W}_l^f\right\|_{2p}^2 \right)^{1/2}, \tag{3.10}$$

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

$$\begin{aligned} C_{X_l} &= [\bar{\psi}_l]_{Lip}^2 \left\| \mathrm{e}^{|Y_l| \vee |\widehat{Y}_l|} \right\|_{2q}^2 + \widetilde{A}_l^2 K_l^2, & C_{W_l^d} &= \widetilde{B}_l^2 K_l^2, \\ C_{Y_l} &= \left(\varphi_d(t_l) \|\psi_{t_l}\|_\infty + [\bar{\psi}_l]_{Lip}\right)^2 \left\| \mathrm{e}^{|Y_l| \vee |\widehat{Y}_l|} \right\|_{2q}^2 + \widetilde{C}_l^2 K_l^2, & C_{W_l^f} &= \widetilde{D}_l^2 K_l^2 \end{aligned} \tag{3.11}$$

*with*

$$K_l = \left\| \mathrm{e}^{|Y_l| \vee |\widehat{Y}_l| + \widetilde{b}_l |W_l^d| \vee |\widehat{W}_l^d|} \right\|_{2q}. \tag{3.12}$$

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

$$\lim_{\bar{N} \to +\infty} \left\|V_k - \widehat{V}_k\right\|_2^2 = 0. \tag{3.13}$$

**Remark 3.8.** From the definition of the processes $X_k$, $W_k^f$, $Y_k$ and $W_k^d$, all are Gaussian random variables hence all the $L^{2q}$-norms in Equations (3.11) and (3.12) are finite. Indeed, let $Z \sim \mathcal{N}(0, \sigma_{_Z})$ a Gaussian random variable with variance $\sigma_{_Z}^2$ and $\widehat{Z}$ an optimal quantizer of $Z$ with cardinality $N$ then $\forall \lambda \in \mathbb{R}_+$

$$\left\| \mathrm{e}^{\lambda |Z| \vee |\widehat{Z}|} \right\|_{2q} = \left( \mathbb{E}\left[ \mathrm{e}^{2q\lambda |Z| \vee |\widehat{Z}|} \right] \right)^{\frac{1}{2q}} \leqslant \left( 2\, \mathbb{E}\left[ \mathrm{e}^{2q\lambda |Z|} \right] \right)^{\frac{1}{2q}} \leqslant 2^{\frac{1}{2q}}\, \mathrm{e}^{q^2 \lambda^2 \sigma_{_Z}^2}. \tag{3.14}$$

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

$$\begin{aligned} |V_k - \widehat{V}_k| \leqslant \max \Big( & \left| h_k(X_k, Y_k) - h_k(\widehat{X}_k, \widehat{Y}_k) \right|, \\ & \left| \mathbb{E}\left[ V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d) \right] - \mathbb{E}\left[ \widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) \right] \right| \Big) \end{aligned} \tag{3.15}$$

thus, using the local Lipschitz property of $h_k$ established in Proposition 2.3 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

$$\begin{aligned} \left\|V_k - \widehat{V}_k\right\|_2^2 &\leqslant \left\|h_k(X_k, Y_k) - h_k(\widehat{X}_k, \widehat{Y}_k)\right\|_2^2 \\ &\qquad + \left\| \mathbb{E}\left[ V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d) \right] - \mathbb{E}\left[ \widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) \right] \right\|_2^2. \\ &\leqslant \left\| \mathrm{e}^{|Y_k| \vee |\widehat{Y}_k|} \right\|_{2q}^2 \left( \left(\varphi_d(t_k) \|\psi_{t_k}\|_\infty + [\bar{\psi}_k]_{Lip}\right)^2 \left\|Y_k - \widehat{Y}_k\right\|_{2p}^2 + [\bar{\psi}_k]_{Lip}^2 \left\|X_k - \widehat{X}_k\right\|_{2p}^2 \right) \\ &\qquad + \left\| \mathbb{E}\left[ V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d) \right] - \mathbb{E}\left[ \widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) \right] \right\|_2^2. \end{aligned} \tag{3.16}$$

Looking at the last term, we have

$$\begin{aligned} \mathbb{E}\left[ V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d) \right] - \mathbb{E}\left[ \widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) \right] \\ = \mathbb{E}\left[ V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d) \right] - \mathbb{E}\left[ V_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) \right] \\ + \mathbb{E}\left[ V_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) \right] - \mathbb{E}\left[ \widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) \right]). \end{aligned} \tag{3.17}$$

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

• For the first term, notice that

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

and

$$\mathbb{E}\left[V_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\right] = Pv_{k+1}(\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) \tag{3.19}$$

then, we directly apply Lemma 2.4 on the function $v_{k+1}$ and obtain

$$\begin{aligned}
&|Pv_{k+1}(X_k, W_k^f, Y_k, W_k^d) - Pv_{k+1}(\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)| \\
&\leqslant \left(\widetilde{A}_k|X_k - \widehat{X}_k| + \widetilde{B}_k|W_k^f - \widehat{W}_k^f| + \widetilde{C}_k|Y_k - \widehat{Y}_k| + \widetilde{D}_k|W_k^d - \widehat{W}_k^d|\right) \mathrm{e}^{|Y_k|\vee|\widehat{Y}_k|+\widetilde{b}_k|W_k^d|\vee|\widehat{W}_k^d|}
\end{aligned} \tag{3.20}$$

with $\widetilde{A}_k$, $\widetilde{B}_k$, $\widetilde{C}_k$, $\widetilde{D}_k$ and $\widetilde{b}_k$ defined by (2.24) and (2.25). Hence, using Hölder's inequality with $p, q \geqslant 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$,

$$\begin{aligned}
&\left\| \mathbb{E}\left[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\right] - \mathbb{E}\left[V_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\right]\right\|_2^2 \\
&\leqslant \left(\widetilde{A}_k^2\|X_k - \widehat{X}_k\|_{2p}^2 + \widetilde{B}_k^2\|W_k^f - \widehat{W}_k^f\|_{2p}^2 + \widetilde{C}_k^2\|Y_k - \widehat{Y}_k\|_{2p}^2 + \widetilde{D}_k^2\|W_k^d - \widehat{W}_k^d\|_{2p}^2\right) \\
&\qquad \times \left\| \mathrm{e}^{|Y_k|\vee|\widehat{Y}_k|+\widetilde{b}_k|W_k^d|\vee|\widehat{W}_k^d|} \right\|_{2q}^2.
\end{aligned} \tag{3.21}$$

• The last one is useful for the induction, indeed

$$\left\| \mathbb{E}\left[V_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\right] - \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\right]\right\|_2^2 \leqslant \left\|V_{k+1} - \widehat{V}_{k+1}\right\|_2^2. \tag{3.22}$$

Finally, using the $L^r$-$L^s$ mismatch theorem for the quadratic optimal quantizers $\widehat{X}_k$ and $\widehat{Y}_k$, if $1 < p < 3/2$, then

$$\begin{aligned}
&\limsup_{N_k^X} N_k^X\|X_k - \widehat{X}_k\|_{2p} < +\infty, \qquad \limsup_{N_k^Y} N_k^Y\|Y_k - \widehat{Y}_k\|_{2p} < +\infty, \\
&\limsup_{N_k^{W^f}} N_k^{W^f}\|W_k^f - \widehat{W}_k^f\|_{2p} < +\infty \quad \text{and} \quad \limsup_{N_k^{W^d}} N_k^{W^d}\|W_k^d - \widehat{W}_k^d\|_{2p} < +\infty
\end{aligned} \tag{3.23}$$

this yields

$$\begin{aligned}
&\left\|V_k - \widehat{V}_k\right\|_2^2 \\
&\leqslant \left\|X_k - \widehat{X}_k\right\|_{2p}^2 \left([\bar{\psi}_k]_{Lip}^2 \left\| \mathrm{e}^{|Y_k|\vee|\widehat{Y}_k|} \right\|_{2q}^2 + \widetilde{A}_k^2 K_k^2\right) \\
&\quad + \left\|Y_k - \widehat{Y}_k\right\|_{2p}^2 \left(\left(\varphi_d(t_k)\|\psi_{t_k}\|_\infty + [\bar{\psi}_k]_{Lip}\right)^2 \left\| \mathrm{e}^{|Y_k|\vee|\widehat{Y}_k|} \right\|_{2q}^2 + \widetilde{C}_k^2 K_k^2\right) \\
&\quad + \widetilde{B}_k^2 K_k^2 \|W_k^f - \widehat{W}_k^f\|_{2p}^2 + \widetilde{D}_k^2 K_k^2 \|W_k^d - \widehat{W}_k^d\|_{2p}^2 + \left\|V_{k+1} - \widehat{V}_{k+1}\right\|_2^2 \\
&\leqslant \sum_{l=k}^{n} C_{X_l}\left\|X_l - \widehat{X}_l\right\|_{2p}^2 + C_{Y_l}\left\|Y_l - \widehat{Y}_l\right\|_{2p}^2 + C_{W_l^d}\left\|W_l^d - \widehat{W}_l^d\right\|_{2p}^2 + C_{W_l^f}\left\|W_l^f - \widehat{W}_l^f\right\|_{2p}^2 \\
&\quad \xrightarrow{\bar{N}\to+\infty} 0
\end{aligned} \tag{3.24}$$

where $K_k = \left\| \mathrm{e}^{|Y_k|\vee|\widehat{Y}_k|+\widetilde{b}_k|W_k^d|\vee|\widehat{W}_k^d|} \right\|_{2q}$ and $\forall k = 1, \ldots, n, \quad C_{X_k}, C_{Y_k}, C_{W_k^d}, C_{W_k^f} < +\infty.$ □

**Remark 3.9.** The same result can be obtained if we relax the assumption on the payoff $\psi_k$. If we only assume the payoff Lipschitz continuous, we have the same limit with the same rate of convergence, however the constants $C_{X_l}, C_{Y_l}, C_{W_l^d}, C_{W_l^f}$ are not the same.

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 $\mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\right]$ are 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 tree and only keep $X$ and $Y$. Doing so, we loose the Markovian property of our original model but we drastically reduce the numerical complexity of the problem. Thence, (2.14) is approximated by

$$\begin{cases} \widehat{V}_n = h_n(\widehat{X}_n, \widehat{Y}_n), \\ \widehat{V}_k = \max\left(h_k(\widehat{X}_k, \widehat{Y}_k), \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\right]\right), \qquad 0 \leqslant k \leqslant n-1 \end{cases} \tag{3.25}$$

where for every $k = 0, \dots, n$, $\widehat{X}_k$ and $\widehat{Y}_k$ are quadratic optimal quantizers of $X_k$ and $Y_k$ of size $N_k^X$ and $N_k^Y$, respectively and we denote $N_k = N_k^X \times N_k^Y$ the size of the grid of the product quantizer.

**Theorem 3.10.** *Let the Markov transition $Pf(x,u,y,v)$ be defined by (2.12) be locally Lipschitz in the sense of Lemma 2.4. Assume that all the payoff functions $(\psi_{t_k})_{k=0:n}$ are Lipschitz continuous with compactly supported (right) derivative. Then the $L^2$-error, induced by the quantization approximation $(\widehat{X}_k, \widehat{Y}_k)$ is upper-bounded by*

$$\begin{aligned} \left\|V_k - \widehat{V}_k\right\|_2 \leqslant \Bigg( \sum_{l=k}^{n-1} & C_{W_{l+1}^f} \left\|W_{l+1}^f - \mathbb{E}[W_{l+1}^f \mid (X_l, Y_l)]\right\|_{2p}^2 + C_{W_{l+1}^d} \left\|W_{l+1}^d - \mathbb{E}[W_{l+1}^d \mid (X_l, Y_l)]\right\|_{2p}^2 \\ & + C_{X_l} \left\|X_l - \widehat{X}_l\right\|_{2p}^2 + C_{Y_l} \left\|Y_l - \widehat{Y}_l\right\|_{2p}^2 \Bigg)^{1/2} \end{aligned} \tag{3.26}$$

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

$$\begin{aligned} C_{X_l} &= [\bar{\psi}_l]_{Lip}^2 \left\| \mathrm{e}^{|Y_l| \vee |\widehat{Y}_l|} \right\|_{2q}^2 + \bar{A}_l^2 \left\| \mathrm{e}^{\bar{b}_l |Y_l| \vee |\widehat{Y}_l|} \right\|_{2q}^2, & C_{W_{l+1}^f} &= B_{l+1}^2 \left\|\widetilde{\kappa}_{k+1}\right\|_{2q}^2, \\ C_{Y_l} &= \left(\varphi_d(t_l) \|\psi_{t_l}\|_\infty + [\bar{\psi}_l]_{Lip}\right)^2 \left\| \mathrm{e}^{|Y_l| \vee |\widehat{Y}_l|} \right\|_{2q}^2 + \bar{C}_l^2 \left\| \mathrm{e}^{\bar{b}_l |Y_l| \vee |\widehat{Y}_l|} \right\|_{2q}^2, & C_{W_{l+1}^d} &= D_{l+1}^2 \left\|\widetilde{\kappa}_{k+1}\right\|_{2q}^2. \end{aligned} \tag{3.27}$$

*Taking the limit in $\bar{N} = \min N_k$, the size of the quadratic optimal quantizers, we have*

$$\lim_{\bar{N} \to +\infty} \left\|V_k - \widehat{V}_k\right\|_2^2 = \sum_{l=k}^{n-1} C_{W_{l+1}^f} \left\|W_{l+1}^f - \mathbb{E}[W_{l+1}^f \mid (X_l, Y_l)]\right\|_{2p}^2 + C_{W_{l+1}^d} \left\|W_{l+1}^d - \mathbb{E}[W_{l+1}^d \mid (X_l, Y_l)]\right\|_{2p}^2. \tag{3.28}$$

*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

$$|V_k - \widehat{V}_k| \leqslant \max\left(\left|h_k(X_k, Y_k) - h_k(\widehat{X}_k, \widehat{Y}_k)\right|, \left|\mathbb{E}\left[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\right] - \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\right]\right|\right) \tag{3.29}$$

thus, using Proposition 2.3 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

$$\begin{aligned}\big\|V_k-\widehat{V}_k\big\|_2^2 &\leqslant \big\|h_k(X_k,Y_k)-h_k(\widehat{X}_k,\widehat{Y}_k)\big\|_2^2 \\ &\qquad + \big\|\,\mathbb{E}\big[V_{k+1}\mid (X_k,W_k^f,Y_k,W_k^d)\big]-\mathbb{E}\big[\widehat{V}_{k+1}\mid(\widehat{X}_k,\widehat{Y}_k)\big]\big\|_2^2 \\ &\leqslant \Big([\bar{\psi}_k]_{Lip}^2\big\|X_k-\widehat{X}_k\big\|_{2p}^2+\big(\varphi_d(t_k)\|\psi_{t_k}\|_\infty+[\bar{\psi}_k]_{Lip}\big)^2\big\|Y_k-\widehat{Y}_k\big\|_{2p}^2\Big)\big\|\,\mathrm{e}^{|Y_k|\vee|\widehat{Y}_k|}\,\big\|_{2q}^2 \\ &\qquad + \big\|\,\mathbb{E}\big[V_{k+1}\mid (X_k,W_k^f,Y_k,W_k^d)\big]-\mathbb{E}\big[\widehat{V}_{k+1}\mid(\widehat{X}_k,\widehat{Y}_k)\big]\big\|_2^2.\end{aligned} \tag{3.30}$$

The last term in Equation (3.30) can be decomposed as follows

$$\begin{aligned}&\mathbb{E}\big[V_{k+1}\mid (X_k,W_k^f,Y_k,W_k^d)\big]-\mathbb{E}\big[\widehat{V}_{k+1}\mid(\widehat{X}_k,\widehat{Y}_k)\big] \\ &\qquad=\mathbb{E}\big[V_{k+1}\mid (X_k,W_k^f,Y_k,W_k^d)\big]-\mathbb{E}\big[V_{k+1}\mid(X_k,Y_k)\big] \\ &\qquad+\mathbb{E}\big[V_{k+1}\mid (X_k,Y_k)\big]-\mathbb{E}\big[V_{k+1}\mid(\widehat{X}_k,\widehat{Y}_k)\big] \\ &\qquad+\mathbb{E}\big[V_{k+1}\mid (\widehat{X}_k,\widehat{Y}_k)\big]-\mathbb{E}\big[\widehat{V}_{k+1}\mid(\widehat{X}_k,\widehat{Y}_k)\big].\end{aligned} \tag{3.31}$$

And again, each term can be upper-bounded.

• The first can be upper-bounded using what we did above on the value function $v_k$ and Hölder's inequality with $p, q \geqslant 1$ such that $\frac{1}{p}+\frac{1}{q}=1$

$$\begin{aligned}&\big\|\,\mathbb{E}\big[V_{k+1}\mid (X_k,W_k^f,Y_k,W_k^d)\big]-\mathbb{E}\big[V_{k+1}\mid(X_k,Y_k)\big]\big\|_2^2 \\ &\quad\leqslant\big\|V_{k+1}-\mathbb{E}\big[V_{k+1}\mid(X_k,Y_k)\big]\big\|_2^2 \\ &\quad\leqslant\big\|v_{k+1}(X_{k+1},W_{k+1}^f,Y_{k+1},W_{k+1}^d) \\ &\qquad\qquad -v_{k+1}\big(X_{k+1},\mathbb{E}\big[W_{k+1}^f\mid(X_k,Y_k)\big],Y_{k+1},\mathbb{E}\big[W_{k+1}^d\mid(X_k,Y_k)\big]\big)\big\|_2^2 \\ &\quad\leqslant\Big\|\Big(B_{k+1}\big|W_{k+1}^f-\mathbb{E}\big[W_{k+1}^f\mid(X_k,Y_k)\big]\big|+D_{k+1}\big|W_{k+1}^d-\mathbb{E}\big[W_{k+1}^d\mid(X_k,Y_k)\big]\big|\Big)\widetilde{\kappa}_{k+1}\Big\|_2^2 \\ &\quad\leqslant\big\|\widetilde{\kappa}_{k+1}\big\|_{2q}^2\Big(B_{k+1}^2\big\|W_{k+1}^f-\mathbb{E}[W_{k+1}^f\mid(X_k,Y_k)]\big\|_{2p}^2+D_{k+1}^2\big\|W_{k+1}^d-\mathbb{E}\big[W_{k+1}^d\mid(X_k,Y_k)\big]\big\|_{2p}^2\Big)\end{aligned} \tag{3.32}$$

with coefficients $b_{k+1}$, $B_{k+1}$ and $D_{k+1}$ defined in (2.31) and (2.32) and

$$\widetilde{\kappa}_{k+1}=\mathrm{e}^{|Y_{k+1}|+b_{k+1}|W_{k+1}^d|\vee|\,\mathbb{E}[W_{k+1}^d|(X_k,Y_k)]|}\,. \tag{3.33}$$

• For the second, we define

$$\widetilde{v}_k(X_k,Y_k)=\mathbb{E}\big[v_{k+1}(X_{k+1},W_{k+1}^f,Y_{k+1},W_{k+1}^d)\mid(X_k,Y_k)\big]. \tag{3.34}$$

Indeed, $\mathbb{E}\big[V_{k+1}\mid(X_k,Y_k)\big]$ is only a function of $X_k$ and $Y_k$, as shown below

$$\begin{aligned}\mathbb{E}\big[V_{k+1}\mid(X_k,Y_k)\big]&=\mathbb{E}\big[v_{k+1}(X_{k+1},W_{k+1}^f,Y_{k+1},W_{k+1}^d)\mid(X_k,Y_k)\big] \\ &=\mathbb{E}\Big[\,\mathbb{E}\big[v_{k+1}(X_{k+1},W_{k+1}^f,Y_{k+1},W_{k+1}^d)\mid(X_k,W_k^f,Y_k,W_k^d)\big]\mid(X_k,Y_k)\Big] \\ &=\mathbb{E}\big[Pv_{k+1}(X_k,W_k^f,Y_k,W_k^d)\mid(X_k,Y_k)\big].\end{aligned} \tag{3.35}$$

Moreover, we can rewrite $W_k^f=\lambda_k X_k \overset{\perp\!\!\!\perp}{+}\xi_k$ and $W_k^d=\widetilde{\lambda}_k Y_k \overset{\perp\!\!\!\perp}{+}\chi_k$ where

$$\lambda_k=\frac{\mathbb{C}\mathrm{ov}(X_k,W_k^f)}{\mathbb{V}\mathrm{ar}(X_k)},\qquad \widetilde{\lambda}_k=\frac{\mathbb{C}\mathrm{ov}(Y_k,W_k^d)}{\mathbb{V}\mathrm{ar}(Y_k)}$$

and $\xi_k \sim \mathcal{N}(0, \sigma^2_{\xi_k})$ and $\chi_k \sim \mathcal{N}(0, \sigma^2_{\chi_k})$ with $\sigma^2_{\xi_k} = \mathbb{V}\mathrm{ar}(W^f_k - \lambda_k X_k)$ and $\sigma^2_{\chi_k} = \mathbb{V}\mathrm{ar}(W^d_k - \widetilde{\lambda}_k Y_k)$, then

$$\begin{aligned}\mathbb{E}\left[Pv_{k+1}(X_k, W^f_k, Y_k, W^d_k) \mid (X_k, Y_k) = (x, y)\right] \\ = \mathbb{E}\left[Pv_{k+1}(x, \lambda_k x + \xi_k, y, \widetilde{\lambda}_k y + \chi_k)\right]\big|_{(x,y)=(X_k,Y_k)}\end{aligned} \tag{3.36}$$

yielding

$$\widetilde{v}_k(x, y) = \mathbb{E}\left[Pv_{k+1}(x, \lambda_k x + \xi_k, y, \widetilde{\lambda}_k y + \chi_k)\right]. \tag{3.37}$$

Now, using Lemma 2.4 on $\widetilde{v}_k$, we have

$$\begin{aligned}&\left|\widetilde{v}_k(x, y) - \widetilde{v}_k(x', y')\right| \\ &= \left|\mathbb{E}\left[Pv_{k+1}(x, \lambda_k x + \xi_k, y, \widetilde{\lambda}_k y + \chi_k) - Pv_{k+1}(x', \lambda_k x' + \xi_k, y', \widetilde{\lambda}_k y' + \chi_k)\right]\right| \\ &\leqslant \mathbb{E}\left[\left|\left((\widetilde{A}_k + \widetilde{B}_k|\lambda_k|)|x - x'| + (1 + \widetilde{C}_k|\widetilde{\lambda}_k|)|y - y'|\right) \mathrm{e}^{(1+\widetilde{b}_k|\widetilde{\lambda}_k|)|y|\vee|y'|+\widetilde{b}_k|\chi_k|}\right|\right] \\ &\leqslant \left(\bar{A}_k|x - x'| + \bar{C}_k|y - y'|\right) \mathrm{e}^{\bar{b}_k|y|\vee|y'|}\end{aligned} \tag{3.38}$$

where

$$\bar{A}_k = (\widetilde{A}_k + \widetilde{B}_k|\lambda_k|)\,\mathbb{E}\left[\,\mathrm{e}^{\widetilde{b}_k|\chi_k|}\,\right], \qquad \bar{C}_k = 1 + \widetilde{C}_k|\widetilde{\lambda}_k|, \tag{3.39}$$

$$\bar{b}_k = 1 + \widetilde{b}_k|\widetilde{\lambda}_k| \tag{3.40}$$

with $\widetilde{A}_k$, $\widetilde{B}_k$, $\widetilde{C}_k$ and $\widetilde{b}_k$ defined in (2.24) and (2.25). Hence, using Hölder's inequality with $p, q \geqslant 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$

$$\begin{aligned}&\left\|\mathbb{E}\left[V_{k+1} \mid (X_k, Y_k)\right] - \mathbb{E}\left[V_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\right]\right\|^2_2 \\ &\qquad = \left\|\widetilde{v}_k(X_k, Y_k) - \widetilde{v}_k(\widehat{X}_k, \widehat{Y}_k)\right\|^2_2 \\ &\qquad \leqslant \left\|\left(\bar{A}_k\left|X_k - \widehat{X}_k\right| + \bar{C}_k\left|Y_k - \widehat{Y}_k\right|\right) \mathrm{e}^{\bar{b}_k|Y_k|\vee|\widehat{Y}_k|}\right\|^2_2 \\ &\qquad \leqslant \left\|\,\mathrm{e}^{\bar{b}_k|Y_k|\vee|\widehat{Y}_k|}\right\|^2_{2q}\left(\bar{A}^2_k\left\|X_k - \widehat{X}_k\right\|^2_{2p} + \bar{C}^2_k\left\|Y_k - \widehat{Y}_k\right\|^2_{2p}\right).\end{aligned} \tag{3.41}$$

- The last one is useful for the induction, indeed

$$\left\|\mathbb{E}\left[V_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\right] - \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\right]\right\|^2_2 \leqslant \left\|V_{k+1} - \widehat{V}_{k+1}\right\|^2_2. \tag{3.42}$$

Finally, using the $L^r$-$L^s$ mismatch theorem on the quadratic optimal quantizers $\widehat{X}_k$ and $\widehat{Y}_k$, if $1 < p < 3/2$, then

$$\limsup_{N^X_k} N^X_k \|X_k - \widehat{X}_k\|_{2p} < +\infty \qquad \text{and} \qquad \limsup_{N^Y_k} N^Y_k \|Y_k - \widehat{Y}_k\|_{2p} < +\infty \tag{3.43}$$

and

$$
\begin{aligned}
&\big\|V_k-\widehat{V}_k\big\|_2^2\\
&\quad\leqslant \big\|X_k-\widehat{X}_k\big\|_{2p}^2\Big([\bar{\psi}_k]_{Lip}^2\big\|\,\mathrm{e}^{|Y_k|\vee|\widehat{Y}_k|}\,\big\|_{2q}^2+\bar{A}_k^2\big\|\,\mathrm{e}^{\bar{b}_k|Y_k|\vee|\widehat{Y}_k|}\,\big\|_{2q}^2\Big)\\
&\qquad+\big\|Y_k-\widehat{Y}_k\big\|_{2p}^2\Big(\big(\varphi_d(t_k)\|\psi_{t_k}\|_\infty+[\bar{\psi}_k]_{Lip}\big)^2\big\|\,\mathrm{e}^{|Y_k|\vee|\widehat{Y}_k|}\,\big\|_{2q}^2+\bar{C}_k^2\big\|\,\mathrm{e}^{\bar{b}_k|Y_k|\vee|\widehat{Y}_k|}\,\big\|_{2q}^2\Big)\\
&\qquad\quad+B_{k+1}^2\big\|\widetilde{\kappa}_{k+1}\big\|_{2q}^2\big\|W_{k+1}^f-\mathbb{E}[W_{k+1}^f\mid(X_k,Y_k)]\big\|_{2p}^2\\
&\qquad\qquad+D_{k+1}^2\big\|\widetilde{\kappa}_{k+1}\big\|_{2q}^2\big\|W_{k+1}^d-\mathbb{E}\left[W_{k+1}^d\mid(X_k,Y_k)\right]\big\|_{2p}^2+\big\|V_{k+1}-\widehat{V}_{k+1}\big\|_2^2\\
&\quad\leqslant\sum_{l=k}^{n-1}C_{W_{l+1}^f}\big\|W_{l+1}^f-\mathbb{E}[W_{l+1}^f\mid(X_l,Y_l)]\big\|_{2p}^2+C_{W_{l+1}^d}\big\|W_{l+1}^d-\mathbb{E}[W_{l+1}^d\mid(X_l,Y_l)]\big\|_{2p}^2\\
&\qquad\qquad+C_{X_l}\big\|X_l-\widehat{X}_l\big\|_{2p}^2+C_{Y_l}\big\|Y_l-\widehat{Y}_l\big\|_{2p}^2\\
&\quad\xrightarrow{\bar{N}\to+\infty}\sum_{l=k}^{n-1}C_{W_{l+1}^f}\big\|W_{l+1}^f-\mathbb{E}[W_{l+1}^f\mid(X_l,Y_l)]\big\|_{2p}^2+C_{W_{l+1}^d}\big\|W_{l+1}^d-\mathbb{E}[W_{l+1}^d\mid(X_l,Y_l)]\big\|_{2p}^2.
\end{aligned}
\tag{3.44}
$$

□

***Practitioner's corner.*** Market implied values of $\sigma_f$, $\sigma_d$ and $\sigma_S$ used for the numerical computations are usually of order

$$
\sigma_f\approx 0.005,\qquad \sigma_d\approx 0.005,\qquad \sigma_S\approx 0.5 \tag{3.45}
$$

and in the most extreme cases, we compute Bermudan options on foreign exchange with maturity 20 years ($T=20$). Thus, we can estimate the order of the induced systematic error. First, we recall the expression of the related coefficients which depends of

$$
\begin{aligned}
B_k&=\sigma_f\frac{T}{n}\sum_{l=k+1}^{n}\max_{l\leqslant i\leqslant n}\left([\bar{\psi}_i]_{Lip}\prod_{j=k+1}^{i}\mathbb{E}[\kappa_j]\right),\\
D_k&=\sigma_d\frac{T}{n}\sum_{l=k+1}^{n}\max_{l\leqslant i\leqslant n}\left((\varphi_d(t_i)\|\psi_i\|_\infty+[\bar{\psi}_i]_{Lip})\prod_{j=k+1}^{i}\mathbb{E}[\kappa_j]\right)
\end{aligned}
\tag{3.46}
$$

with

$$
\kappa_j=\mathrm{e}^{|G_j^3|+b_j|G_j^4|},\qquad \widetilde{\kappa}_{l+1}=\mathrm{e}^{|Y_{l+1}|+b_{l+1}|W_{l+1}^d|\vee|\,\mathbb{E}[W_{l+1}^d|(X_l,Y_l)]|} \tag{3.47}
$$

and

$$
b_k=\sigma_d T\left(1-\frac{k-1}{n}\right). \tag{3.48}
$$

Now, considering the case where the payoffs are the same at each exercise date, the Lipschitz constants can be upper-bounded by $[\bar{\psi}]_{Lip}$:

$$
[\bar{\psi}_k]_{Lip}=[\psi_{t_k}]_{Lip}S_0\varphi_f(t_k)\,\mathrm{e}^{-\sigma_S^2t_k/2}\,\|\psi'_{t_k}\|_\infty\,\mathrm{e}^c\leqslant S_0[\psi_{t_k}]_{Lip}\|\psi'_{t_k}\|_\infty\,\mathrm{e}^c=:[\bar{\psi}]_{Lip} \tag{3.49}
$$

and let $\kappa$ defined by

$$
\kappa=\max_k\mathbb{E}[\kappa_k]=\mathbb{E}\left[\,\mathrm{e}^{|G_0^3|+b_0|G_0^4|}\,\right]\leqslant\frac{1}{2}\,\mathbb{E}\left[\,\mathrm{e}^{2|G_0^3|}+\mathrm{e}^{2b_0|G_0^4|}\,\right] \tag{3.50}
$$

moreover, if $Z\sim\mathcal{N}(0,\sigma^2)$ then $\mathbb{E}\left[\,\mathrm{e}^{\lambda|Z|}\,\right]=\mathrm{e}^{\lambda^2\sigma^2/2}$, thence we can upper-bound $\kappa$

$$
\kappa\leqslant\frac{1}{2}\,\mathbb{E}\left[\,\mathrm{e}^{\sigma_3^2/2}+\mathrm{e}^{b_0^2/2}\,\right]=\frac{1}{2}\,\mathbb{E}\left[\,\mathrm{e}^{\sigma_d^2/96}+\mathrm{e}^{\sigma_d^2T^2/2}\,\right]\approx 1. \tag{3.51}
$$

$\kappa$ being bounded, we notice that the main constants $B_k^2$ and $D_k^2$ in the remaining error are of order $\sigma_d^2$ or $\sigma_f^2$, indeed

$$\begin{aligned} B_k &\leqslant \sigma_f \frac{T}{n} [\bar{\psi}]_{Lip} (n-k)\kappa^{n-k} \approx \sigma_f \frac{T}{n} [\bar{\psi}]_{Lip} (n-k), \\ D_k &\leqslant \sigma_d \frac{T}{n} \big( \max_l \varphi_d(t_l) \|\psi\|_\infty + [\bar{\psi}]_{Lip} \big)(n-k)\kappa^{n-k} \approx \sigma_d \frac{T}{n} (\|\psi\|_\infty + [\bar{\psi}]_{Lip})(n-k). \end{aligned} \tag{3.52}$$

Furthermore

$$\begin{aligned} \mathbb{E}\left[\widetilde{\kappa}_{k+1}^{2q}\right] &= \mathbb{E}\left[ \mathrm{e}^{2q|Y_{k+1}| + 2qb_{k+1}|W_{k+1}^d| \vee |\,\mathbb{E}[W_{k+1}^d|(X_k,Y_k)]|} \right] \\ &\leqslant \frac{1}{2}\left( \mathbb{E}\left[ \mathrm{e}^{4q|Y_{k+1}|} \right] + \mathbb{E}\left[ \mathrm{e}^{4qb_{k+1}|W_{k+1}^d| \vee |\,\mathbb{E}[W_{k+1}^d|(X_k,Y_k)]|} \right] \right) \\ &\leqslant \frac{1}{2}\left( \mathbb{E}\left[ \mathrm{e}^{4q|Y_{k+1}|} \right] + \mathbb{E}\left[ \mathrm{e}^{4q\sigma_d(T-t_k)|W_{k+1}^d| \vee |\,\mathbb{E}[W_{k+1}^d|(X_k,Y_k)]|} \right] \right) \\ &\leqslant \frac{1}{2}\left( \mathrm{e}^{8q^2\sigma_d^2 T^3/3} + 2\,\mathrm{e}^{8q^2\sigma_d^2 (T-t_k)^2 t_{k+1}} \right) \end{aligned} \tag{3.53}$$

and from elementary inequality $(a+b)^{1/q} \leqslant a^{1/q} + b^{1/q}$, $a, b \geqslant 0$, $q \geqslant 1$

$$\begin{aligned} \left\|\widetilde{\kappa}_{k+1}\right\|_{2q}^2 = \mathbb{E}\left[\widetilde{\kappa}_{k+1}^{2q}\right]^{\frac{1}{q}} &\leqslant \left( \frac{1}{2}\,\mathrm{e}^{8q^2\sigma_d^2 T^3/3} + \mathrm{e}^{8q^2\sigma_d^2 (T-t_k)^2 t_{k+1}} \right)^{\frac{1}{q}} \\ &\leqslant \left( \frac{1}{2}\,\mathrm{e}^{8q^2\sigma_d^2 T^3/3} \right)^{\frac{1}{q}} + \left( \mathrm{e}^{8q^2\sigma_d^2 (T-t_k)^2 t_{k+1}} \right)^{\frac{1}{q}} \\ &\leqslant \frac{1}{2^{1/q}}\,\mathrm{e}^{8q\sigma_d^2 T^3/3} + \mathrm{e}^{8q\sigma_d^2 (T-t_k)^2 t_{k+1}}\,. \end{aligned} \tag{3.54}$$

The two terms on the right-hand side of the inequality do not explode. Indeed, the function $g : t \mapsto (T-t)^2 t$, defined for $t \in [0,T]$ with $T = 20$, attains its maximum on $t = 20/3$ and $g(20/3) \approx 1185$, hence for the considered values

$$\forall k = 1, \ldots, n, \qquad \left\|\widetilde{\kappa}_{k+1}\right\|_{2q}^2 \leqslant C_{\widetilde{\kappa}} \approx 6. \tag{3.55}$$

Finally, rewriting the obtained systematic error induced by the approximation with this new

informations in (3.28) we have

$$
\begin{aligned}
\big\|V_k - \widehat{V}_k\big\|_2^2 \xrightarrow{N\to+\infty} &\sum_{l=k}^{n-1} B_{l+1}^2 \big\|\widetilde{\kappa}_{l+1}\big\|_{2q}^2 \big\|W_{l+1}^d - \mathbb{E}[W_{l+1}^d \mid (X_l, Y_l)]\big\|_{2p}^2 \\
&\qquad + D_{l+1}^2 \big\|\widetilde{\kappa}_{l+1}\big\|_{2q}^2 \big\|W_{l+1}^f - \mathbb{E}[W_{l+1}^f \mid (X_l, Y_l)]\big\|_{2p}^2 \\
\leqslant &\sigma_f^2 \Big(\frac{T}{n}\Big)^2 [\bar{\psi}]_{Lip}^2 \sum_{l=k}^{n-1} (n-l)^2 \kappa^{2(n-l)} C_{\widetilde{\kappa}} \big\|W_{l+1}^d - \mathbb{E}[W_{l+1}^d \mid (X_l, Y_l)]\big\|_{2p}^2 \\
&+ \sigma_d^2 \Big(\frac{T}{n}\Big)^2 \big(\max_l \varphi_d(t_l) \|\psi\|_\infty + [\bar{\psi}]_{Lip}\big)^2 \\
&\qquad \times \sum_{l=k}^{n-1} (n-l)^2 \kappa^{2(n-l)} C_{\widetilde{\kappa}} \big\|W_{l+1}^f - \mathbb{E}[W_{l+1}^f \mid (X_l, Y_l)]\big\|_{2p}^2 \\
\leqslant &2\sigma_f^2 \Big(\frac{T}{n}\Big)^2 [\bar{\psi}]_{Lip}^2 \sum_{l=k}^{n-1} (n-l)^2 \kappa^{2(n-l)} C_{\widetilde{\kappa}} \big\|W_{l+1}^d\big\|_{2p}^2 \\
&+ 2\sigma_d^2 \Big(\frac{T}{n}\Big)^2 \big(\max_l \varphi_d(t_l) \|\psi\|_\infty + [\bar{\psi}]_{Lip}\big)^2 \sum_{l=k}^{n-1} (n-l)^2 \kappa^{2(n-l)} C_{\widetilde{\kappa}} \big\|W_{l+1}^f\big\|_{2p}^2 \\
\leqslant &\Big(\sigma_f^2 [\bar{\psi}]_{Lip}^2 + \sigma_d^2 \big(\max_l \varphi_d(t_l) \|\psi\|_\infty + [\bar{\psi}]_{Lip}\big)^2\Big) 4 \frac{C_{\widetilde{\kappa}}}{\pi^{1/3}} \Big(\frac{T}{n}\Big)^2 \sum_{l=k}^{n-1} t_{l+1} (n-l)^2 \kappa^{2(n-l)}.
\end{aligned}
\tag{3.56}
$$

Hence, the systematic error is upper-bounded by the squared volatilities $\sigma_d^2$ and $\sigma_f^2$. These parameters being of order $5 \times 10^{-3}$ at most, the systematic error is negligible as long as these volatilities stay reasonably small.

**Remark 3.11.** As in the Markov case, we can extend this result to the case where the payoffs $(\psi_k)_k$ are Lipschitz continuous, however the residual error can not be as easily estimated and controlled.

# 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 3.10 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 time is 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 time 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 4.1.** The computation times given below measure the time needed for loading the pre-computed 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.

**Characterisation 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 $N_k = N$, $k = 1 \dots, n$ 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 $\sigma_S$ compare to $\sigma_d$. 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.** Let $(x_{i_1}^k)_{i_1=1:N^X}$, $(u_{i_2}^k)_{i_2=1:N^{W^f}}$, $(y_{i_3}^k)_{i_3=1:N^Y}$ and $(v_{i_4}^k)_{i_4=1:N^{W^d}}$ be the associated centroids of $\widehat{X}_k$, $\widehat{W}_k^f$, $\widehat{Y}_k$ and $\widehat{W}_k^d$ respectively, at a given time $t_k$ with $0 \leqslant k \leqslant n$. Using the discrete property of the optimal quantizers, the conditional expectation appearing in (3.9) can be rewritten as

$$\begin{aligned}
&\mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) = \left(x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k\right)\right] \\
&\quad = \mathbb{E}\left[\widehat{v}_{k+1}(\widehat{X}_{k+1}, \widehat{W}_{k+1}^f, \widehat{Y}_{k+1}, \widehat{W}_{k+1}^d) \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) = \left(x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k\right)\right] \\
&\quad = \sum_{j_1,j_2,j_3,j_4} \pi_{i,j}^{(\mathrm{M}),k}\, \widehat{v}_{k+1}\left(x_{j_1}^{k+1}, u_{j_2}^{k+1}, y_{j_3}^{k+1}, v_{j_4}^{k+1}\right)
\end{aligned} \tag{4.1}$$

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

$$\begin{aligned}
\pi_{i,j}^{(\mathrm{M}),k} = \mathbb{P}\Big(&\left(\widehat{X}_{k+1}, \widehat{W}_{k+1}^f, \widehat{Y}_{k+1}, \widehat{W}_{k+1}^d\right) = \left(x_{j_1}^{k+1}, u_{j_2}^{k+1}, y_{j_3}^{k+1}, v_{j_4}^{k+1}\right) \\
&\mid \left(\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d\right) = \left(x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k\right)\Big).
\end{aligned}$$

Due to the dimension of the problem (4 in this case), we cannot compute these probabilities 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 $\widetilde{\pi}_{i,j}^{(\mathrm{M}),k}$, where the conditional part $\left\{\left(\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d\right) = \left(x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k\right)\right\}$ is replaced by $\left\{(X_k, W_k^f, Y_k, W_k^d) = \left(x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k\right)\right\}$, yielding

$$\begin{aligned}
\widetilde{\pi}_{i,j}^{(\mathrm{M}),k} = \mathbb{P}\Big(&\left(\widehat{X}_{k+1}, \widehat{W}_{k+1}^f, \widehat{Y}_{k+1}, \widehat{W}_{k+1}^d\right) = \left(x_{j_1}^{k+1}, u_{j_2}^{k+1}, y_{j_3}^{k+1}, v_{j_4}^{k+1}\right) \\
&\mid (X_k, W_k^f, Y_k, W_k^d) = \left(x_{i_1}^k, y_{i_2}^k, u_{i_3}^k, v_{i_4}^k\right)\Big).
\end{aligned} \tag{4.2}$$

The reason for replacing $\{(\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) = (x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k)\}$ by $\{(X_k, W_k^f, Y_k, W_k^d) = (x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k)\}$ is explained in the next paragraph dealing with the Non-Markovian case with lighter notations (see Equation (4.5) and (4.7)). 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.

These remarks allow us to rewritte the QBDPP in the Markovian case (3.9) as

$$\begin{cases} \widehat{v}_n(x_{i_1}^n, u_{i_2}^n, y_{i_3}^n, v_{i_4}^n) = h_n(x_{i_1}^n, y_{i_3}^n), \\ \widehat{v}_k(x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k) = \max\left(h_k(x_{i_1}^k, y_{i_3}^k), \sum_{j_1,j_2,j_3,j_4} \widetilde{\pi}_{i,j}^{(\mathrm{M}),k}\, \widehat{v}_{k+1}(x_{j_1}^{k+1}, u_{j_2}^{k+1}, y_{j_3}^{k+1}, v_{j_4}^{k+1})\right). \end{cases} \tag{4.3}$$

**The algorithm: Non-Markovian case.** Let $(x_{i_1}^k)_{i_1=1:N^X}$ and $(y_{i_3}^k)_{i_3=1:N^Y}$ be the associated centroids of $\widehat{X}_k$ and $\widehat{Y}_k$ respectively, at a given time $t_k$ with $0 \leqslant k \leqslant n$. Again, as in the Markovian case, using the discrete property of the optimal quantizers, the conditional expectation appearing in (3.25) can be rewritten as

$$\begin{aligned} \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k) = (x_{i_1}^k, y_{i_2}^k)\right] &= \mathbb{E}\left[\widehat{v}_{k+1}(\widehat{X}_{k+1}, \widehat{Y}_{k+1}) \mid (\widehat{X}_k, \widehat{Y}_k) = (x_{i_1}^k, y_{i_2}^k)\right] \\ &= \sum_{j_1,j_2} \pi_{i,j}^{(\mathrm{NM}),k}\, \widehat{v}_{k+1}(x_{j_1}^{k+1}, y_{j_2}^{k+1}) \end{aligned} \tag{4.4}$$

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

$$\pi_{i,j}^{(\mathrm{NM}),k} = \mathbb{P}\left((\widehat{X}_{k+1}, \widehat{Y}_{k+1}) = (x_{j_1}^{k+1}, y_{j_2}^{k+1}) \mid (\widehat{X}_k, \widehat{Y}_k) = (x_{i_1}^k, y_{i_2}^k)\right).$$

This probability can be computed by numerical integration, ie

$$\begin{aligned} \pi_{i,j}^{(\mathrm{NM}),k} &= \mathbb{P}\left((\widehat{X}_{k+1}, \widehat{Y}_{k+1}) = (x_{j_1}^{k+1}, y_{j_2}^{k+1}) \mid (\widehat{X}_k, \widehat{Y}_k) = (x_{i_1}^k, y_{i_2}^k)\right) \\ &= \mathbb{P}\left((\widehat{X}_{k+1}, \widehat{Y}_{k+1}) = (x_{j_1}^{k+1}, y_{j_2}^{k+1}) \mid X_k \in (x_{i_1-1/2}^k, x_{i_1+1/2}^k), Y_k \in (y_{i_2-1/2}^k, y_{i_2+1/2}^k)\right) \\ &= \int_{x_{i_1-1/2}^k}^{x_{i_1+1/2}^k} \int_{y_{i_2-1/2}^k}^{y_{i_2+1/2}^k} \mathbb{P}\left((\widehat{X}_{k+1}, \widehat{Y}_{k+1}) = (x_{j_1}^{k+1}, y_{j_2}^{k+1}) \mid (X_k, Y_k) = (x, y)\right) f_\Sigma(x, y)dx\, dy \end{aligned} \tag{4.5}$$

where $f_\Sigma(x, y)$ is the joint density of a centered bivariate Gaussian vector with covariance matrix $\Sigma$ given by

$$\Sigma = \begin{pmatrix} \mathbb{V}\mathrm{ar}(X_k) & \mathbb{C}\mathrm{ov}(X_k, Y_k) \\ \mathbb{C}\mathrm{ov}(X_k, Y_k) & \mathbb{V}\mathrm{ar}(Y_k) \end{pmatrix}. \tag{4.6}$$

However, computing the probability in Equation (4.5) can be too time consuming, hence once again, we approximate this probability by $\widetilde{\pi}_{i,j}^{(\mathrm{NM}),k}$, 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

$$\widetilde{\pi}_{i,j}^{(\mathrm{NM}),k} = \mathbb{P}\left((\widehat{X}_{k+1}, \widehat{Y}_{k+1}) = (x_{j_1}^{k+1}, y_{j_2}^{k+1}) \mid (X_k, Y_k) = (x_{i_1}^k, y_{i_2}^k)\right). \tag{4.7}$$

From the definition of an optimal quantizer and Equation (2.9), this probability can be rewritten as the probability that a correlated bivariate normal distribution lies in a rectangular

domain

$$\begin{aligned}\widetilde{\pi}_{i,j}^{(\mathrm{NM}),k} &= \mathbb{P}\left(\widehat{X}_{k+1} = x_{j_1}^{k+1}, \widehat{Y}_{k+1} = y_{j_2}^{k+1} \mid X_k = x_{i_1}^k, Y_k = y_{i_2}^k\right)\\ &= \mathbb{P}\left(X_{k+1} \in \left(x_{j_1-1/2}^{k+1}, x_{j_1+1/2}^{k+1}\right), Y_{k+1} \in \left(y_{j_2-1/2}^{k+1}, y_{j_2+1/2}^{k+1}\right) \mid X_k = x_{i_1}^k, Y_k = y_{i_2}^k\right)\\ &= \mathbb{P}\left(x_{i_1}^k + \sigma_f \delta W_k^f + G_{k+1}^1 \in \left(x_{j_1-1/2}^{k+1}, x_{j_1+1/2}^{k+1}\right), y_{i_2}^k - \sigma_d \delta W_k^d + G_{k+1}^3 \in \left(y_{j_2-1/2}^{k+1}, y_{j_2+1/2}^{k+1}\right)\right)\\ &= \mathbb{P}\left(Z^1 \in \left(x_{j_1-1/2}^{k+1} - x_{i_1}^k, x_{j_1+1/2}^{k+1} - x_{i_1}^k\right), Z^2 \in \left(y_{j_2-1/2}^{k+1} - y_{i_2}^k, y_{j_2+1/2}^{k+1} - y_{i_2}^k\right)\right)\end{aligned} \tag{4.8}$$

where

$$\begin{pmatrix} Z^1 \\ Z^2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{Z^1}^2 & \rho_{Z^1,Z^2}\sigma_{Z^1}\sigma_{Z^2} \\ \rho_{Z^1,Z^2}\sigma_{Z^1}\sigma_{Z^2} & \sigma_{Z^2}^2 \end{pmatrix}\right) \tag{4.9}$$

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

The advantage of expressing (4.8) 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.

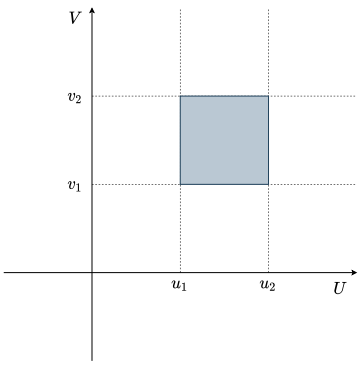



Figure 2

Indeed, let $(U, V)$ a two-dimensional correlated and standardized normal distribution with correlation $\rho$ 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]. In our case, we are interested in the computation of probabilities of the form

$$\mathbb{P}\left(U \in (u_1, u_2), V \in (v_1, v_2)\right). \tag{4.10}$$

This probability is represented graphically as the integral of the two-dimensional density over the rectangular domain in grey in Figure 2. Now, using $F_{U,V}^{\rho}(u, v)$, the probability (4.10) is given by

$$\mathbb{P}\left(U \in (u_1, u_2), V \in (v_1, v_2)\right) = F_{U,V}^{\rho}(u_2, v_2) - F_{U,V}^{\rho}(u_1, v_2) - F_{U,V}^{\rho}(u_2, v_1) + F_{U,V}^{\rho}(u_1, v_1). \tag{4.11}$$

This remark will allow us to reduce drastically the computation time induced by the evaluation of the conditional probabilities and so, of the conditional expectations.

Now, going back to our problem, the QBDPP in the non-Markovian case rewrites (3.25)

$$\begin{cases} \widehat{v}_n\left(x_{i_1}^n, y_{i_2}^n\right) = h_n\left(x_{i_1}^n, y_{i_2}^n\right), \qquad 1 \leqslant i_1 \leqslant N_n^X, \qquad 1 \leqslant i_2 \leqslant N_n^Y, \\ \widehat{v}_k\left(x_{i_1}^k, y_{i_2}^k\right) = \max\left(h_k\left(x_{i_1}^k, y_{i_2}^k\right), \sum_{j_1, j_2} \pi_{i,j}^{(\mathrm{NM}),k}\, \widehat{v}_{k+1}\left(x_{j_1}^{k+1}, y_{j_2}^{k+1}\right)\right). \end{cases} \tag{4.12}$$

In order to test numerically the two methods, we will evaluate PRDC European and Bermudan options with maturities $2Y$, $5Y$ and $10Y$. 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 $\sigma_d$ and $\sigma_f$.

| | | | | | |
|---|---|---|---|---|---|
| $P_d(0,t)$ | $\exp(-r_d t)$ | $r_d$ | 0.015 | $\rho_{Sd}$ | 0 |
| $P_f(0,t)$ | $\exp(-r_f t)$ | $r_f$ | 0.01 | $\rho_{Sf}$ | 0 |
| $S_0$ | 88.17 | $\sigma_S$ | 0.5 | $\rho_{df}$ | 0 |

Table 1: *Market values.*

| | | | |
|---|---|---|---|
| $\forall k \in 1,\dots,n, \quad C_d(t_k)$ | 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 4.2.** When the correlations $\rho_{df}$ and $\rho_{Sd}$ are equal to zero, the numerical computation of probabilities $\widetilde{\pi}_{i,j}^{(\mathrm{M}),k}$ and $\widetilde{\pi}_{i,j}^{(\mathrm{NM}),k}$ can be accelerated. Indeed, in the Markovian case, (4.2) can be rewritten as

$$\begin{aligned}\widetilde{\pi}_{i,j}^{(\mathrm{M}),k} = \mathbb{P}\left((\widehat{X}_{k+1}, \widehat{W}^f_{k+1}) = \left(x^{k+1}_{j_1}, u^{k+1}_{j_2}\right) \mid (X_k, W^f_k) = \left(x^k_{i_1}, u^k_{i_2}\right)\right) \\ \times \mathbb{P}\left((\widehat{Y}_{k+1}, \widehat{W}^d_{k+1}) = \left(y^{k+1}_{j_3}, v^{k+1}_{j_4}\right) \mid (Y_k, W^d_k) = \left(y^k_{i_3}, v^k_{i_4}\right)\right).\end{aligned} \tag{4.13}$$

In that case, we can use the CDF of a correlated bivariate normal distribution, as detailed above for the non-Markovian case in (4.11), for computing these probabilities in a very effective and faster way rather than performing a Monte Carlo simulation.

In the non-Markovian case, (4.8) can be rewritten as

$$\begin{aligned}\widetilde{\pi}_{i,j}^{(\mathrm{NM}),k} &= \mathbb{P}\left(Z^1 \in \left(x^{k+1}_{j_1-1/2} - x^k_{i_1}, x^{k+1}_{j_1+1/2} - x^k_{i_1}\right)\right) \mathbb{P}\left(Z^2 \in \left(y^{k+1}_{j_2-1/2} - y^k_{i_2}, y^{k+1}_{j_2+1/2} - y^k_{i_2}\right)\right) \\ &= \left(F_Z\Big(\frac{x^{k+1}_{j_1+1/2} - x^k_{i_1}}{\sigma_{Z^1}}\Big) - F_Z\Big(\frac{x^{k+1}_{j_1-1/2} - x^k_{i_1}}{\sigma_{Z^1}}\Big)\right)\left(F_Z\Big(\frac{y^{k+1}_{j_2+1/2} - y^k_{i_2}}{\sigma_{Z^2}}\Big) - F_Z\Big(\frac{y^{k+1}_{j_2-1/2} - y^k_{i_2}}{\sigma_{Z^2}}\Big)\right)\end{aligned} \tag{4.14}$$

where $F_Z(\cdot)$ is the CDF of a one-dimensional normal distribution, $\sigma_{Z^1}$ is the standard deviation of $Z^1$ and $\sigma_{Z^2}$ 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

First of all, we compare the asymptotic behaviour of the Markovian and the non-Markovian approaches when pricing European PRDC Options with different volatilities and maturities. In this case, we consider only two dates in the tree: $t_0 = 0$ an $t_n = T$, the algorithm is a regular cubature formula and no systematic error is induced by the non-markovianity of the couple $(X_k, Y_k)$. These numerical tests confirm that both approaches give the same value, however the non-Markovian approach converges much faster due to the dimension of the product quantization, 2 for the first one and 4 for the last one. Indeed, the complexity of the 2 dimensional pricer is of order of $N = N^X \times N^Y$ while the complexity of the 4 dimensional pricer is of order $N = N^X \times N^Y \times N^{W^d} \times N^{W^f}$. $N$ being the size of the product quantizer at each date (in two dimensions: $N = N^X \times N^Y$ and in four dimensions $N = N^X \times N^{W^f} \times N^Y \times N^{W^d}$).

In the case of the European options, we have a closed-form formula for the price of (2.4). The benchmark price is computed using the rewriting of (2.4) as a sum of Calls: at a time $t_k$,

the payoff can be expressed as

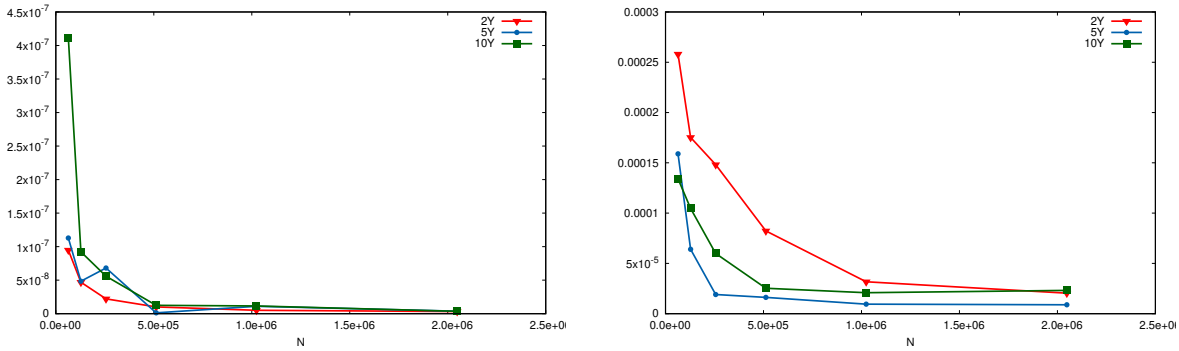
$$\begin{aligned}\psi_{t_k}(S_{t_k}) &= \min\left(\max\left(\frac{C_f(t_k)}{S_0}S_{t_k} - C_d(t_k), \text{Floor}(t_k)\right), \text{Cap}(t_k)\right)\\ &= \text{Floor}(t_k) - a_k(S_{t_k} - K_k^1)_+ + a_k(S_{t_k} - K_k^2)_+\end{aligned}$$

with $a_k = \dfrac{C_f(t_k)}{S_0}$, $K_k^1 = \dfrac{\text{Cap}(t_k) + C_d(t_k)}{C_f(t_k)} \times S_0$ and $K_k^2 = \dfrac{\text{Floor}(t_k) + C_d(t_k)}{C_f(t_k)} \times S_0$ and the closed-form formula for the price of a Call is detailed in Appendix B. The prices given by the closed-form formula of the European options we consider (different values of volatilities and different maturities) are given in Table 3.

| | Exact price | |
|---|---|---|
| $T$ \ $\sigma$ | 50*bp* | 500*bp* |
| 2Y | 2.171945242 | 2.159404007 |
| 5Y | 1.630435483 | 1.539295559 |
| 10Y | 1.127330259 | 0.8013151892 |

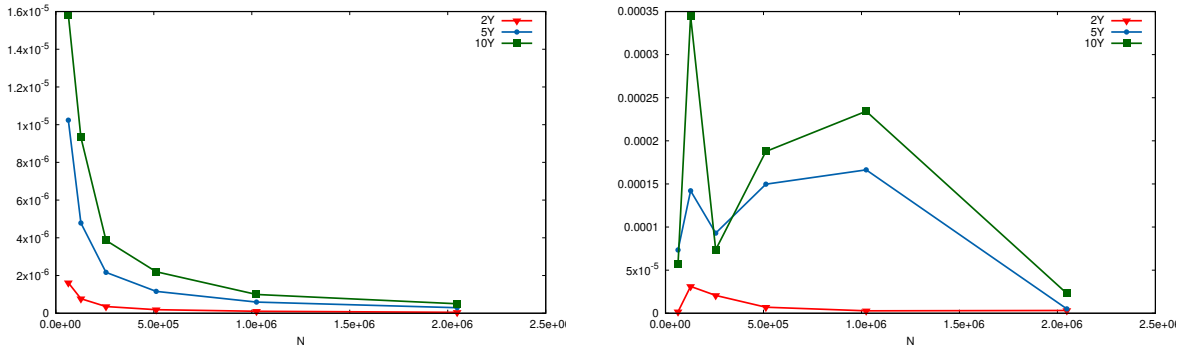
Table 3: *Prices given by closed-form formula of European options with zero correlations.* ($\sigma_d = \sigma_f = \sigma$)

The difference of speed of convergence between the two methods is illustrated in Figures 3 and 4 for the relative errors for both methods compared to the benchmark. $N$ in the label of each graphic represents the size of the product quantizer ($N^X \times N^{W^f} \times N^Y \times N^{W^d}$ in the Markovian case and $N^X \times N^Y$ in the other case), hence the complexity of both trees are the same.




(a) *Non-Markovian – 2d* (b) *Markovian – 4d*

Figure 3: $\sigma_d = \sigma_f = 50bp$ *– Relative errors for both methods for 2Y, 5Y and 10Y European options pricing (with zero correlations).*




(a) *Non-Markovian – 2d* (b) *Markovian – 4d*

Figure 4: $\sigma_d = \sigma_f = 500bp$ – *Relative errors for both methods for 2Y, 5Y and 10Y European options pricing (with zero correlations).*

For both methods, a relative error of $1bp$ is quickly reached, even for high values of $\sigma_d$ and $\sigma_f$. Indeed, the time needed in order to achieve a $1bp$ precision for building a quantization tree with 2 dates, computing the probabilities and pricing a European option is at most 6 ms for the non-Markovian method and at most 85ms for the Markovian one when the correlations are equal to zero. The computation times needed for a $1bp$ relative error are summarised in Table 4.

| | Non-Markovian – 2d | | Markovian – 4d | |
|---|---|---|---|---|
| $T$ \ $\sigma$ | $50bp$ | $500bp$ | $50bp$ | $500bp$ |
| 2Y | 1 ms (32000) | 4 ms (32000) | 24 ms (512000) | 4 ms (64000) |
| 5Y | 4 ms (32000) | 6 ms (32000) | 4 ms (64000) | 85 ms (2048000) |
| 10Y | 4 ms (32000) | 3 ms (32000) | 14 ms (256000) | 83 ms (2048000) |

Table 4: *Times in milliseconds needed for reaching a* $1bp$ *precision for European options pricing with zero correlations using both methods with, in parenthesis, the size* $N$ *of the grid at each time step.* $(\sigma_d = \sigma_f = \sigma)$

**Remark 4.3.** Of course, the pricers can be used even when we consider non-zero correlations. We choose to show only the asymptotic behaviour of the non-Markovian method because it converges much faster and the computations of the probabilities can be made deterministically using the CDF of a correlated bivariate normal distribution. However, if we want to use the Markovian approach, we need to compute the transition probabilities using a Monte Carlo simulation, which is a drawback for the method as it increases its computation time. We consider the following correlations

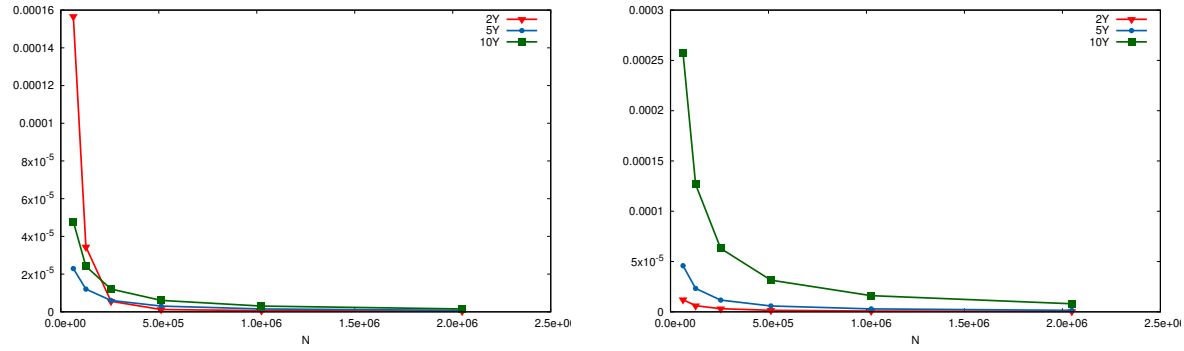
$$\rho_{Sf} = -0.0272, \qquad \rho_{Sd} = 0.1574, \qquad \rho_{df} = 0.6558.$$

Table 5 summarises the prices given by the closed-form formula.

| | Exact price | |
|---|---|---|
| $T$ \ $\sigma$ | $50bp$ | $500bp$ |
| 2Y | 2.173803852 | 2.185536786 |
| 5Y | 1.636518082 | 1.652226813 |
| 10Y | 1.141944391 | 1.103531914 |

Table 5: *Prices given by closed-form formula of European options with correlations.* ($\sigma_d = \sigma_f = \sigma$)

Figures 5a and 5b display the relative error induced by the numerical method as a function of $N$. And in Table 6, we summarise the computation needed in order to reach a $1bp$ relative error.




(a) $\sigma_d = \sigma_f = 50bp$ (b) $\sigma_d = \sigma_f = 500bp$

Figure 5: *Relative errors for the non-Markovian method for 2Y, 5Y and 10Y European options pricing (with correlations).*

| | Non-Markovian – 2d | |
|---|---|---|
| $T$ \ $\sigma$ | $50bp$ | $500bp$ |
| 2Y | 71 ms (64000) | 34 ms (32000) |
| 5Y | 31 ms (32000) | 31 ms (32000) |
| 10Y | 32 ms (32000) | 139 ms (128000) |

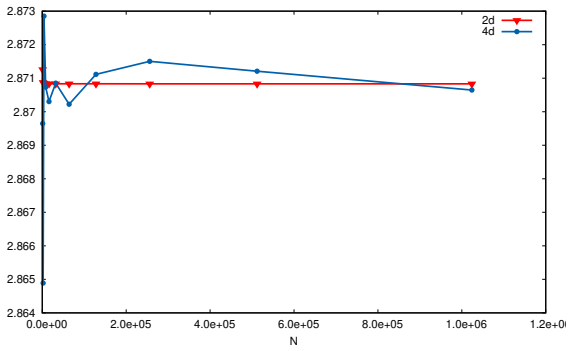
Table 6: *Times in milliseconds needed for reaching a $1bp$ relative error of European options pricing with correlations using the non-Markovian method with, in parenthesis, the size $N$ of the grid at each time step.* ($\sigma_d = \sigma_f = \sigma$)

It is clear that one should prefer the non-Markovian methodology to the Markovian one for the evaluation of European options as it is a fast and accurate method for producing prices in the 3-factor model.

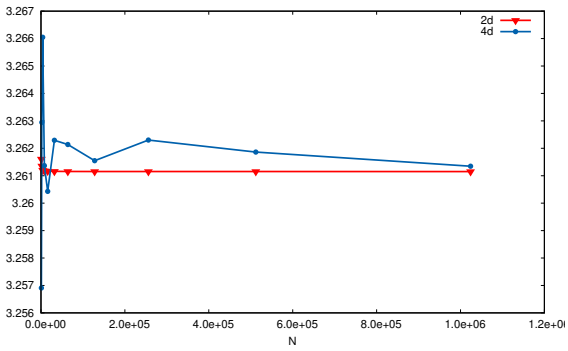
## 4.2 Bermudan option

Now, we compare the asymptotic behaviour of both approaches when pricing true 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 and 10 years) and the domestic/foreign volatilities ($50bp$ and $500bp$).

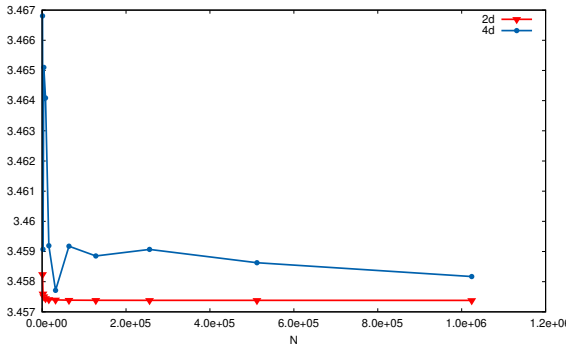
When using domestic and foreign volatilities close to market values, we observe numerically that the non-Markovian method converges a lot faster than the Markovian one for a given complexity. However both methods do not converge to the same value (see Figures 6a, 6b, 6c), which is consistent with the results we found in Theorems 3.7 and 3.10. As in the European case, $N$ in the label of each graph represents the size of the product quantizer ($N^X \times N^{W^f} \times N^Y \times N^{W^d}$ in the Markovian case and $N^X \times N^Y$ in the other case), hence the complexity of both trees are the same.




(a) *2Y*




(b) *5Y*




(c) *10Y*

Figure 6: $\sigma_d = \sigma_f = 50bp$ – *Price with the two methods for 2Y, 5Y and 10Y yearly exercisable Bermudan options (with zero correlations).*

However, the relative systematic error induced by the non-Markovian methodology is negligible as can be seen in Figure 7, at most $5bp$ for a 10-year annual Bermudan option. Hence, one should prefer, again, the non-Markovian methodology when considering to evaluate Bermudan options.

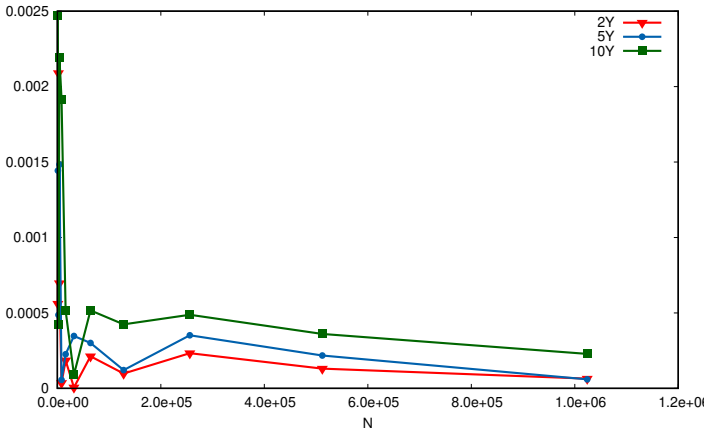



Figure 7: $\sigma_d = \sigma_f = 50bp$ – *Relative differences between the two methods for 2Y, 5Y and 10Y yearly exercisable Bermudan options (with zero correlations).*

**Remark 4.4.** If we consider more exercise dates for the Bermudan option, the systematic errors increase, as shown in Figure 8 where we considered Bermudan options exercisable every 6 months and the same parameters as before with zero correlations and $\sigma_d = \sigma_f = 50bp$. However, even-though the error is higher for small $N$, when the non-Markovian pricer has converged, the relative difference between both methods is still acceptable (lower than $5bp$).

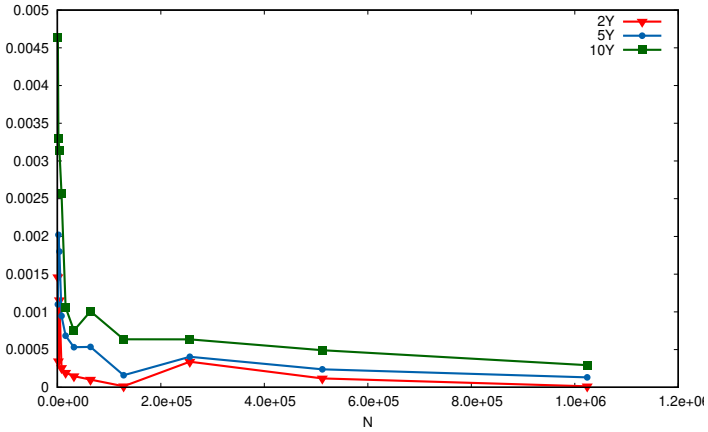
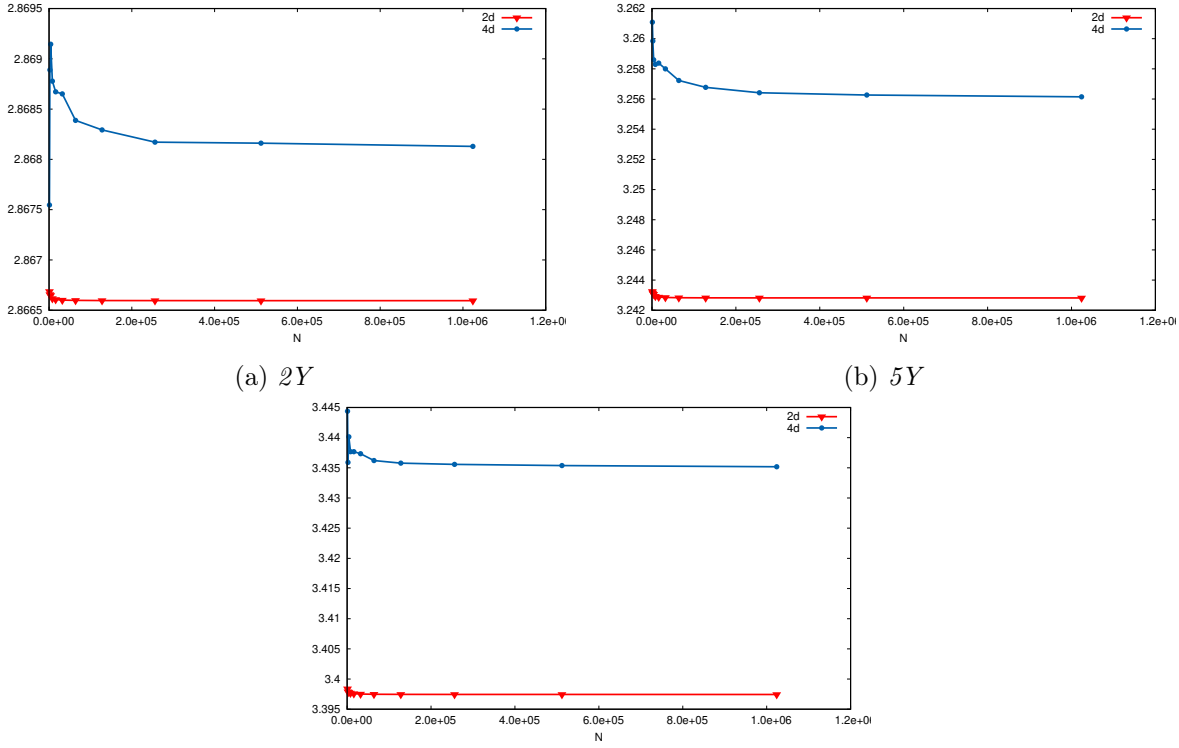



Figure 8: $\sigma_d = \sigma_f = 50bp$ – *Relative differences between the two methods for 2Y, 5Y and 10Y bi-annual exercisable Bermudan options (with zero correlations).*

When we consider higher values the volatilities, $\sigma_d = \sigma_f = 500bp$, as expected the prices produced by the non-Markovian methodology produce a systematic error bigger than the case where $\sigma_d = \sigma_f = 50bp$ (see Figures 9a, 9b, 9c and 10). However, the relative difference between the two methods after convergence is reasonable: less than 0.1% for expiry 2 years, 0.4% for 5 years and around 1.1% for 10 years.




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

(c) *10Y*

Figure 9: $\sigma_d = \sigma_f = 500bp$ – *Price with the two methods for 2Y, 5Y and 10Y yearly exercisable Bermudan options (with zero correlations).*

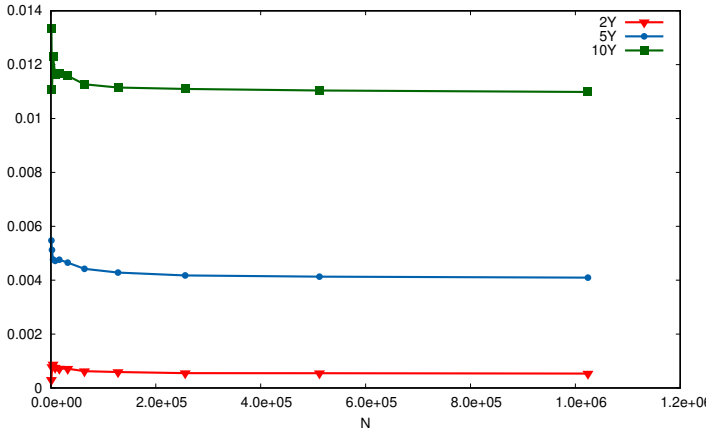



Figure 10: $\sigma_d = \sigma_f = 500bp$ – *Relative differences between the two methods for 2Y, 5Y and 10Y yearly exercisable Bermudan options (with zero correlations).*

In Figure 7, we reference the time needed for reaching a $5bp$ relative precision (we compare the price given by grids of size $N$ to the "asymptotic", which is the price given by the same method with a very large $N$) for the pricing of Bermudan options in a scenario of zero correlations. The non-Markovian method attains better precision than a relative precision of $5bp$ in a few milliseconds, at most 7 ms where the Markovian one can need 4 seconds for reaching that precision. Hence, the 2 dimensional approximation seems again to be the better choice.

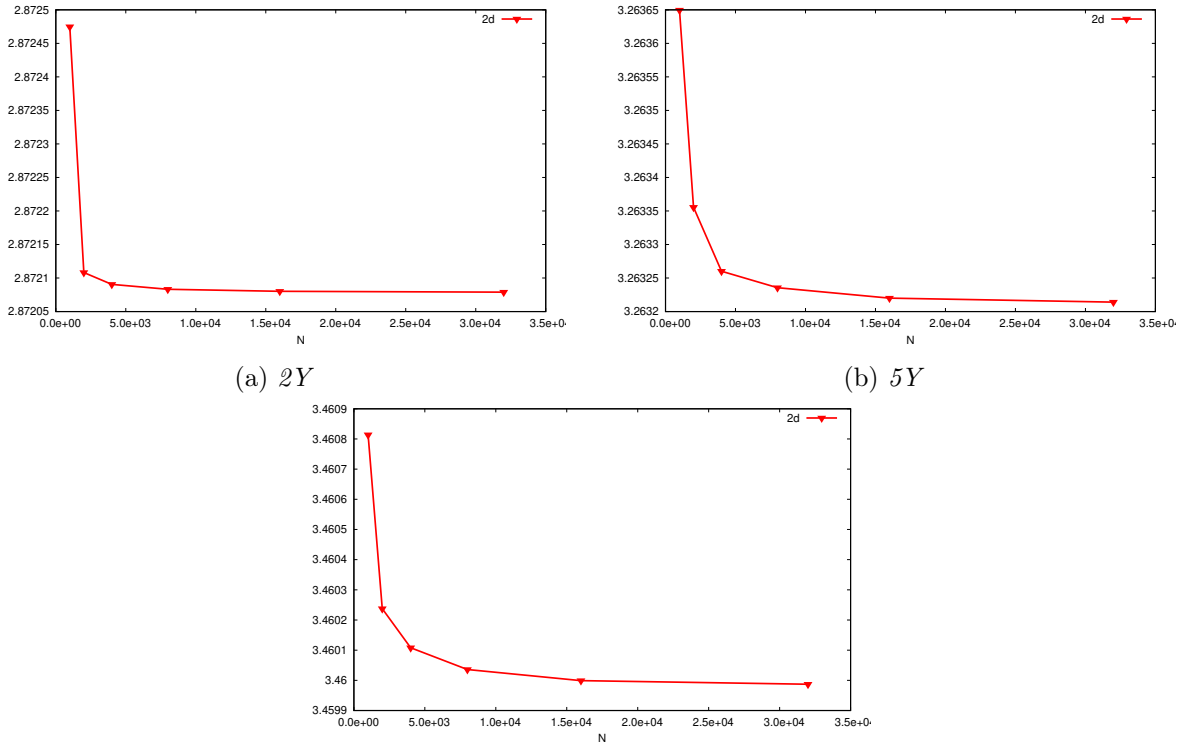
| $T$ \ $\sigma$ | Non-Markovian – 2d | | Markovian – 4d | |
|---|---|---|---|---|
| | $50bp$ | $500bp$ | $50bp$ | $500bp$ |
| 2Y | 1 ms (1000) | 1 ms (1000) | 25 ms (8000) | 4 ms (1000) |
| 5Y | 3 ms (1000) | 4 ms (1000) | 98 ms (8000) | 1903 ms (64000) |
| 10Y | 7 ms (1000) | 7 ms (1000) | 468 ms (16000) | 3850 ms (64000) |

Table 7: *Times in milliseconds needed for reaching a* $5bp$ *relative precision for Bermudan options pricing using both methods with zero correlation and, in parenthesis, the size* $N$ *of the grid at each time step.* ($\sigma_d = \sigma_f = \sigma$)

**Remark 4.5.** Again, the pricers can even be used when we consider non-zero correlations and we choose to show only the asymptotic behaviour of the non-Markovian method, for the same reasons as the European case. We consider the same correlations as in the European case

$$\rho_{Sf} = -0.0272, \qquad \rho_{Sd} = 0.1574, \qquad \rho_{df} = 0.6558.$$

Figures 11a, 11b and 11c display the price given by the numerical method as a function of $N$ and Table 8 summarises the computation time needed in order to do better than a $3bp$ precision.




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

(c) *10Y*

Figure 11: $\sigma_d = \sigma_f = 50bp$ – *Price of 2Y, 5Y and 10Y yearly exercisable Bermudan options using the non-Markovian method (with correlations).*

| | Non-Markovian – 2d |
|---|---|
| $T$ \ $\sigma$ | $50bp$ |
| 2Y | 122 ms (1000) |
| 5Y | 553 ms (1000) |
| 10Y | 1283 ms (1000) |

Table 8: *Times in milliseconds needed for reaching a 3bp relative precision for Bermudan yearly exercisable options pricing with correlations using the non-Markovian method with, in parenthesis, the size $N$ of the grid at each time step. ($\sigma_d = \sigma_f = \sigma$)*

## 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 author reports no conflicts of interest. The author alone is 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

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

or equivalently

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

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 $\forall t > s$

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

the last equality is ensured if and only if

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

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 $V_0$

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

**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*[2]

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

*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

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

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

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

We focus on the first term

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

We do the following change of numéraire:

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

with

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

[2]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.

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 $\widetilde{\mathbb{Q}}$:

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

and $S_t$ becomes

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

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

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

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

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

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 $X \sim \text{Log-}\mathcal{N}(\mu, \sigma^2)$, 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 $X \overset{\tilde{\mathbb{Q}}}{\sim} \text{Log-}\mathcal{N}(-\mu(0,t), \sigma^2(0,t))$, 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: $\sigma_S(s) = \sigma_S$, $\sigma_d(s,t) = \sigma_d \times (t-s)$ $\sigma_f(s,t) = \sigma_f \times (t-s)$

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

□