



**HAL**  
open science

# Computing XVA for American basket derivatives by Machine Learning techniques

Ludovic Goudenège, Andrea Molent, Antonino Zanette

► **To cite this version:**

Ludovic Goudenège, Andrea Molent, Antonino Zanette. Computing XVA for American basket derivatives by Machine Learning techniques. 2022. hal-04308564

**HAL Id: hal-04308564**

**<https://hal.science/hal-04308564>**

Preprint submitted on 27 Nov 2023

**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.

# Computing XVA for American basket derivatives by Machine Learning techniques

LUDOVIC GOUDENEGE\*

ANDREA MOLENT †

ANTONINO ZANETTE‡

---

## Abstract

Total value adjustment (XVA) is the change in value to be added to the price of a derivative to account for the bilateral default risk and the funding costs. In this paper, we compute such a premium for American basket derivatives whose payoff depends on multiple underlyings. In particular, in our model, those underlyings are supposed to follow the multidimensional Black-Scholes stochastic model. In order to determine the XVA, we follow the approach introduced by Burgard and Kjaer [9] and afterward applied by Arregui et al. [2, 3] for the one-dimensional American derivatives. The evaluation of the XVA for basket derivatives is particularly challenging as the presence of several underlyings leads to a high-dimensional control problem. We tackle such an obstacle by resorting to Gaussian Process Regression, a machine learning technique that allows one to address the curse of dimensionality effectively. Moreover, the use of numerical techniques, such as control variates, turns out to be a powerful tool to improve the accuracy of the proposed methods. The paper includes the results of several numerical experiments that confirm the goodness of the proposed methodologies.

*Keywords:* XVA; Gaussian Process Regression; Basket option; Control variates

---

---

\*Fédération de Mathématiques de CentraleSupélec - CNRS FR3487, France - [ludovic.goudege@math.cnrs.fr](mailto:ludovic.goudege@math.cnrs.fr)

†Dipartimento di Scienze Economiche e Statistiche, Università degli Studi di Udine, Italy - [andrea.molent@uniud.it](mailto:andrea.molent@uniud.it)

‡Dipartimento di Scienze Economiche e Statistiche, Università degli Studi di Udine, Italy - [antonino.zanette@uniud.it](mailto:antonino.zanette@uniud.it)

# 1 Introduction

After the financial crisis of 2007 and the default of several financial institutions, practitioners, regulators, and finally, academics have turned increasing attention to counterparty risk. Currently, careful and weighted management of counterparty risk is required at the legislative level by the Basel III agreements of 2010, as well as codified by the IFRS standard starting from 2013. Consequently, when assessing the value of an OTC derivative instrument, banks must apply a series of adjustments to the risk-free price, capable of accounting for the costs associated with the effects of a possible default of any of the two counterparties. The entirety of these corrections is known as the total credit value adjustment, usually indicated by the abbreviation XVA. The main elements that contribute to the calculation of the XVA are the CVA, the DVA and the FVA. The CVA, credit value adjustment, is the premium that an agent must charge to the counterparty to cover the losses that could derive from the default of the same. In particular, these losses occur when the value of the contract is positive for the agent and the counterparty, following a default, is unable to comply with the contractual terms. The DVA, debit value adjustment, is the consideration of the CVA for the counterparty: in the event of the bankruptcy of the agent, he is no longer obliged to comply with the contractual responsibilities and, if the derivative has a negative value for the agent, then he draws a benefit, to the detriment of the counterparty. Finally, the FVA, funding value adjustment, is the change in value in the derivative that comes from the costs or benefits, which the agent obtains following the collateralization of the contract.

Recently, these issues have attracted the attention of many academics and nowadays the literature on credit value adjustment is large. The most common approach to XVA valuation consists in computing the price of the contract subject to risk through a PDE. One of the first authors to suggest a PDE based approach is Piterbarg [23], who introduces a model to include funding costs on derivative valuations when collateral has to be posted. Burgard and Kjaer [9] propose a more general model for the evaluation of bilateral counterparty risk and funding costs still based on the description of the value of European-type derivatives in terms of PDE. De Graaf et al. [11, 12] propose the so-called finite-difference Monte Carlo (FDMC) method, which exploits both finite-difference and Monte Carlo methods to compute the CVA and to compute first and second-order sensitivities for counterparty credit risk. Feng [15] adapts the FDMC method to deal with the case of an underlying evolving according to the Bates model: in this particular case, the PDE to be solved is replaced by a partial integral differential equation (PIDE), which implies an additional computational effort. Goudenège et al. [17] improve the method proposed by De Graaf et al. and compute the CVA in the Bates model by solving coupled PIDEs.

Other authors have considered the Monte Carlo method. Ballotta et al. [4] use Monte Carlo and Fourier transform based methods to study a structural model when the underlying follows a Lévy process. Brigo and Vrins [8] use Monte Carlo to evaluate CVA in a model that effectively manages wrong-way risk. Antonelli et al. [1] propose a procedure based on a Taylor approximation for evaluating XVA and compare it against Monte Carlo simulations.

Recently, Arregui et al. [2, 3] extend the model of Burgard and Kjaer [9] to the analysis of American-type derivatives. This line of research is taken up by Salvador and Oosterlee [26], who develop the stochastic model for the underlying by considering stochastic volatility. Furthermore, Yuan et al. [29] present two different numerical approaches to estimate the total value adjustments of the Bermudan option, under the pure jump CGMY model.

Numerical techniques for option pricing that rely solely on PDEs generally suffer from the curse of dimensionality, that is the explosion of computational cost in the presence of high dimensional problems, whereas standard Monte Carlo methods are not effective in the case of American options. The previously

discussed methods for calculating XVA are not exempt from this limitation.

Newer techniques for evaluating derivatives make use of machine learning methods. In this regard, some authors employ neural networks. For example, Lapeyre and Lelong [20] study the Longstaff and Schwartz algorithm when the standard least-square regression is replaced by a neural network approximation. Becker et al. [5, 6, 7] develop deep learning methods for pricing and hedging American-style and, more generally, for solving optimal control problems. Other authors exploit Gaussian Process Regression (GPR), a Machine Learning technique that allows for estimations from scattered data in large dimensional spaces. In this regard, we mention the work of Ludkovski [22], who evaluates Bermudan options by fitting the continuation values through GPR. More recently, Goudenège et al. [18] propose three GPR-based algorithm, termed GPR-MC, GPR-Tree and GPR-EI, for pricing American options on a basket of assets following multi-dimensional Black-Scholes dynamics.

The literature on the computation of the XVA for high dimensional derivatives is rather sparse. As far as the computation of the credit adjustments are concerned, She and Greu [28] compute CVA and DVA by employing neural network as a universal approximator. Crépey and Dixon [10] exploit GPR to speed up the computation of the CVA for derivatives portfolios. Gnoatto et al. [16] exploit artificial neural network to compute the XVA for large portfolios of derivatives. Despite the importance of this topic, to our knowledge, no one has ever studied the calculation of XVA for American basket options, which are probably the most popular option involving several assets.

In this paper, we aim to fill this gap, by proposing an approach based on a suitable probabilistic formulation of the XVA, derived from the model of Burgard and Kjaer [9], which exploits the GPR-MC and the GPR-EI algorithms for option pricing to overcome the curse of dimensionality. We point out that we have chosen to consider Burgard and Kjaer’s model as it is particularly suitable as, unlike other models, the American option exercise strategy is shaped to take into account the probability of default of any of the agents. Moreover, depending on the choice of the mark-to market value, two possible kinds of models are considered: a linear and a non-linear. Furthermore, the computation accuracy is increased by exploiting suitable control variate for both the riskless and the risky price. Numerous numerical tests demonstrate the reliability and accuracy of the proposed procedures when different derivatives are considered.

The remainder of the paper is organized as follows. In Section 2 we introduce the model for XVA on American basket options. In Section 3 we describe the proposed procedures. In Section 4 we discuss numerical results. Finally, in Section 5, we conclude.

## 2 Total value adjustment for American basket options

In this Section, we describe a PDE-based model for the total value adjustment when American options are concerned and we discuss a probabilistic interpretation that we are going to exploit for our approaches. We stress out that the model we develop here is inspired by the framework previously introduced by Burgard and Kjaer [9] and developed by Arregui et al. [2, 3], which is very interesting among the others because it allows the exercise strategy of the American option to be influenced by the possibility of default of each agent. This phenomenon, which is certainly plausible in reality, is not present in other models. For example, the model by De Graaf et al. [12, 11], which is usually employed by other authors, considers the strategy for the risky option to be the same as the strategy for options without default risk but, in our opinion, this does not seem to be the right choice. This aspect is pointed out in the following Remark.

*Remark 1. Consider an American call option, which is at the money at the time of issue. Now, suppose an agent buys such an option from a counterparty that provides a null recovery rate and that is going to*

default before the maturity of the option almost sure. It is well known that if the underlying does not pay dividends, and the counterparty does not default, it is never optimal to exercise an American call option before maturity. So, if the agent employs the standard strategy, he will achieve a payoff equal to zero almost sure, as the default of the counterparty will occur before maturity and the option will lose all its value (the recovery rate is zero). On the other hand, if he exercises the option immediately, he will obtain a positive payoff (the option is in-the-money), so this strategy is better than the classical one. This simple example shows that the optimal strategy for exercising an American option must take the default risk into account.

Let

$$\mathbf{S} = (\mathbf{S}_t)_{t \in [0, T]} = (S_t^1, \dots, S_t^d)_{t \in [0, T]}$$

denote a  $d$ -dimensional stochastic process following the multi-dimensional Black-Scholes model. Under the risk neutral probability  $\mathbb{Q}$ , the dynamics of each underlying is given by

$$dS_t^i = (r - \eta_i) S_t^i dt + \sigma_i S_t^i dW_t^i, \quad i = 1, \dots, d, \quad (2.1)$$

with  $\mathbf{S}_0 = (s_0^1, \dots, s_0^d) \in \mathbb{R}_+^d$  the spot price,  $r$  the (constant) interest rate,  $\eta = (\eta_1, \dots, \eta_d)$  the vector of dividend rates,  $\sigma = (\sigma_1, \dots, \sigma_d)$  the vector of volatilities,  $\mathbf{W}$  a  $d$ -dimensional correlated Brownian motion and  $\rho_{ij}$  the instantaneous correlation coefficient between  $W_t^i$  and  $W_t^j$ .

Let us consider an American option issued at time 0 with maturity  $T$  and let  $H : \mathbb{R}_+^d \rightarrow \mathbb{R}$  denote the payoff function. Let us term B the issuer and C the buyer of the option. For the moment, we suppose that none of the two agents can default. We approximate the value of the risk-less American option by a Bermudan option which can be exercise at the times  $t_n = n \cdot \Delta t$  for  $n = 0, \dots, N$  with  $\Delta t = T/N$  and  $N \in \mathbb{N}$ . By employing standard arguments, one can prove that

$$V(t_n, \mathbf{S}_t) = \max(C(t, \mathbf{S}_t), H(\mathbf{S}_t)),$$

where  $C(t, \mathbf{S}_t)$  stands for the continuation value. In particular,  $C(t, \mathbf{S}_t)$  restricted to the time interval  $]t_n, t_{n+1}[$  is equal to  $C^n(t, \mathbf{S}_t)$ , the solution of the following PDE, defined in  $]t_n, t_{n+1}[$  for  $n = 0, \dots, N - 1$  :

$$\frac{\partial C^n}{\partial t} + \mathcal{A}(C^n) - rC^n = 0,$$

with  $C^n = C^n(t_{n+1}, \mathbf{x})$  for  $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_d)$  and

$$\mathcal{A}(C^n) = \sum_{i=1}^d (r - \eta_i) \mathbf{x}_i \frac{\partial C^n}{\partial \mathbf{x}_i} + \sum_{i=1}^d \frac{\sigma_i^2 \mathbf{x}_i^2}{2} \frac{\partial^2 C^n}{\partial \mathbf{x}_i^2} + \sum_{i=1}^{d-1} \sum_{j=i+1}^d \rho_{i,j} \sigma_i \sigma_j \mathbf{x}_i \mathbf{x}_j \frac{\partial^2 C^n}{\partial \mathbf{x}_i \partial \mathbf{x}_j}.$$

The terminal condition is

$$C^n(t_{n+1}, \mathbf{x}) = \begin{cases} H(\mathbf{x}) & \text{if } n + 1 = N, \\ V(t_{n+1}, \mathbf{x}) & \text{otherwise.} \end{cases}$$

Now, let us suppose that both agents B and C can default. We take the point of B, and we denote the risky option price by  $\hat{V}(t, \mathbf{S}_t, J_t^B, J_t^C)$ , with  $J^B$  and  $J^C$  two independent jump processes that change value, from 0 to 1, at the time the corresponding agent defaults.

Let  $M_t = M(t, \mathbf{S}_t)$  represent the close-out mark-to-market value, that is, the monetary value of the contract used as the basis for settlement. Let us define  $M^+ = \max(M, 0)$  and  $M^- = \min(M, 0)$ . Following Burgard and Kjaer [9], in case of default of one counterparty, the risky values are defined as follows:

- if the issuer B defaults first,

$$\hat{V}(t, \mathbf{S}_t, 1, 0) = M_t^+ + R_B M_t^-,$$

with  $R_B \in [0, 1]$  the recovery rate of C respect to the default of B;

- if the buyer C defaults first,

$$\hat{V}(t, \mathbf{S}_t, 0, 1) = R_C M_t^+ + M_t^-,$$

with  $R_C \in [0, 1]$  the recovery rate of B respect to the default of C.

Let  $\lambda_B$  and  $\lambda_C$  be the constant default intensities of B and C, respectively, and  $s_F$  the funding cost of B. According to Burgard and Kjaer [9], if the derivative can be used as a collateral, then  $s_F = 0$ , and if it cannot, then  $s_F = (1 - R_B) \lambda_B$ . Following Arregui et al. [2, 3], the value  $\hat{V}(t, \mathbf{S}_t, 0, 0)$  of the Bermudan risky option, satisfies

$$\hat{V}(t, \mathbf{S}_t, 0, 0) = \max\left(\hat{C}(t, \mathbf{S}_t), H(\mathbf{S}_t)\right),$$

with  $\hat{C}$  the continuation value of the risky option. Similarly to what happens for the risk-free option,  $\hat{C}(t, \mathbf{S}_t)$  restricted to the time interval  $]t_n, t_{n+1}[$  is equal to  $\hat{C}^n(t, \mathbf{S}_t)$ , the solution of the following PDE, defined in  $]t_n, t_{n+1}[$  for  $n = 0, \dots, N - 1$ :

$$\frac{\partial \hat{C}^n}{\partial t} + \mathcal{A}(\hat{C}^n) - r\hat{C}^n = (\lambda_B + \lambda_C) \hat{C}^n + s_F M^+ - \lambda_B (R_B M^- + M^+) - \lambda_C (R_C M^+ + M^-), \quad (2.2)$$

with the terminal condition

$$\hat{C}^n(t_{n+1}, \mathbf{x}) = \begin{cases} H(\mathbf{x}) & \text{if } n + 1 = N, \\ \hat{V}(t_{n+1}, \mathbf{x}, 0, 0) & \text{otherwise.} \end{cases}$$

We proceed backward in time. Suppose  $\hat{V}^n(t_{n+1}, \mathbf{x})$  is known and we aim to compute  $\hat{V}^n(t_n, \mathbf{x})$ . By the Feynman-Kac formula applied to equation (2.2), (see e.g. Platen and Heath [24]), we have

$$\hat{C}^n(t_n, \mathbf{x}) = \mathbb{E}^{\mathbb{Q}} \left[ \int_{t_n}^{t_{n+1}} e^{-r_0(u-t_n)} g(u, \mathbf{S}_u) du + e^{-r_0 \Delta t} \hat{C}^n(t_{n+1}, \mathbf{S}_{t_{n+1}}) \mid S_{t_n} = \mathbf{x} \right], \quad (2.3)$$

with

$$r_0 = r + \lambda_B + \lambda_C,$$

$$\begin{aligned} g(u, \mathbf{S}_u) &= - [s_F M_u^+ - \lambda_B (R_B M_u^- + M_u^+) - \lambda_C (R_C M_u^+ + M_u^-)] \\ &= M_u^+ (\lambda_B + \lambda_C R_C - s_F) + M_u^- (\lambda_C + \lambda_B R_B) \\ &= M_u^+ c_p + M_u^- c_m, \end{aligned}$$

$c_p = \lambda_B + \lambda_C R_C - s_F$  and  $c_m = \lambda_C + \lambda_B R_B$ . In particular, as  $\lambda_B, \lambda_C, R_C$  and  $R_B$  are positive quantities and  $s_F$  is equal to 0 or  $(1 - R_B) \lambda_B$ , thus  $c_p$  and  $c_m$  are positive values.

We approximate the integral in (2.3) by a two points trapezoidal quadrature rule:

$$\begin{aligned} \hat{C}^n(t_n, \mathbf{x}) &\approx \mathbb{E}^{\mathbb{Q}} \left[ \frac{e^{-r_0 \Delta t} g(t_{n+1}, \mathbf{S}_{t_{n+1}}) + g(t_n, \mathbf{S}_{t_n})}{2} \Delta t + e^{-r_0 \Delta t} \hat{C}^n(t_{n+1}, \mathbf{S}_{t_{n+1}}) \mid S_{t_n} = \mathbf{x} \right] \\ &= e^{-r_0 \Delta t} \mathbb{E}^{\mathbb{Q}} \left[ \frac{\Delta t}{2} g(t_{n+1}, \mathbf{S}_{t_{n+1}}) + \hat{C}^n(t_{n+1}, \mathbf{S}_{t_{n+1}}) \mid S_{t_n} = \mathbf{x} \right] + \frac{\Delta t}{2} g(t_n, \mathbf{S}_{t_n}), \end{aligned}$$

and thus

$$\hat{V}(t_n, \mathbf{x}, 0, 0) \approx \max \left\{ e^{-r_0 \Delta t} \mathbb{E}^{\mathbb{Q}} \left[ \frac{\Delta t}{2} g(t_{n+1}, \mathbf{S}_{t_{n+1}}) + \hat{V}(t_{n+1}, \mathbf{S}_{t_{n+1}}, 0, 0) \mid S_{t_n} = \mathbf{x} \right] + \frac{\Delta t}{2} g(t_n, \mathbf{x}), H(\mathbf{x}) \right\} \quad (2.4)$$

Now, we distinguish two cases:  $M_u = V(u, S_u)$ , that is the value of the risk-free derivative, and  $M_u = \hat{V}(u, S_u, 0, 0)$ , that is the value of the defaultable derivative.

## 2.1 Case $M = V$

We suppose that the values of  $V$  have already been computed in a suitable domain. If  $M_u = V(u, \mathbf{S}_u)$ , we can compute  $\hat{V}(t_n, \mathbf{S}_{t_n}, 0, 0)$  explicitly, by replacing  $M$  with the pre-computed values of  $V$  and by approximating the expectation in (2.4) by a suitable numeric technique.

## 2.2 Case $M = \hat{V}$

If  $M_u = \hat{V}(u, S_u)$ , then

$$\hat{V}(t_n, \mathbf{x}, 0, 0) \approx \max \left\{ E(\mathbf{x}) + \frac{\Delta t}{2} \left( \hat{V}(t_n, \mathbf{x}, 0, 0)^+ c_p + \hat{V}(t_n, \mathbf{x}, 0, 0)^- c_m \right), H(\mathbf{x}) \right\}, \quad (2.5)$$

with

$$E(\mathbf{x}) = e^{-r_0 \Delta t} \mathbb{E}^{\mathbb{Q}} \left[ \frac{\Delta t}{2} \left( \hat{V}(t_{n+1}, \mathbf{S}_{t_{n+1}}, 0, 0)^+ c_p + \hat{V}(t_{n+1}, \mathbf{S}_{t_{n+1}}, 0, 0)^- c_m \right) + \hat{V}(t_{n+1}, \mathbf{S}_{t_{n+1}}, 0, 0) \mid S_{t_n} = \mathbf{x} \right]. \quad (2.6)$$

We define  $\tilde{V}(t_n, \mathbf{x})$  as the solution of the implicit equation problem

$$\tilde{V}(t_n, \mathbf{x}) = \max \left\{ E(\mathbf{x}) + \frac{\Delta t}{2} \left( \tilde{V}(t_n, \mathbf{x})^+ c_p + \tilde{V}(t_n, \mathbf{x})^- c_m \right), H(\mathbf{x}) \right\}, \quad (2.7)$$

and we employ it as an approximation of  $\hat{V}(t_n, \mathbf{x}, 0, 0)$ . Equation (2.7) is implicit –  $\hat{V}$  appears both on left and the right side of the equation – and non linear. The following proposition discuss how to solve it.

*Proposition 2.1.* Let  $\tilde{V}(t_n, \mathbf{x})$  be the unique solution of the implicit equation (2.7). Then, if  $H(\mathbf{x}) \leq 0$ :

- if  $E(\mathbf{x}) \leq H(\mathbf{x}) (1 - \frac{\Delta t}{2} c_m) \leq 0$  then  $\tilde{V}(t_n, \mathbf{x}) = H(\mathbf{x})$ ;
- if  $H(\mathbf{x}) (1 - \frac{\Delta t}{2} c_m) < E(\mathbf{x}) \leq 0$  then  $\tilde{V}(t_n, \mathbf{x}) = \frac{E(\mathbf{x})}{1 - \frac{\Delta t}{2} c_m}$ ;
- if  $E(\mathbf{x}) > 0$  then  $\tilde{V}(t_n, \mathbf{x}) = \frac{E(\mathbf{x})}{1 - \frac{\Delta t}{2} c_p}$ .

If  $H(\mathbf{x}) > 0$ :

- if  $E(\mathbf{x}) \leq H(\mathbf{x}) (1 - \frac{\Delta t}{2} c_p)$  then  $\tilde{V}(t_n, \mathbf{x}) = H(\mathbf{x})$ ;
- if  $E(\mathbf{x}) > H(\mathbf{x}) (1 - \frac{\Delta t}{2} c_p)$  then  $\tilde{V}(t_n, \mathbf{x}) = \frac{E(\mathbf{x})}{1 - \frac{\Delta t}{2} c_p}$ .

The proof of Proposition 2.1 is discussed in the Appendix A.

## 3 Gaussian Process Regression for computing XVA

According to the previous Section, the calculation of XVA requires the computation of an expected value, both in the case  $M = V$  and in the case  $M = \hat{V}$ , see (2.4). This calculation involves a stochastic underlying which is a multidimensional process, potentially high dimensional. We propose to use two techniques, already successfully applied by Goudenège et al. [18] for multidimensional option pricing problems: GPR-MC and GPR-EI.

Below, we recall the main aspects of these two methods, and we refer the interested reader to [18] for more information.

### 3.1 GPR-MC

The GPR Monte Carlo approach employs Monte Carlo simulations to compute the continuation value of a Bermudan option and GPR to learn the option value at each time step.

The algorithm starts by simulating a set of trajectories of the underlyings. Let  $X^n$  represent the set of  $P$  points whose coordinates represent certain possible values for the underlyings at time  $t_n$ , for  $n = 0, \dots, N$ , that is

$$X^n = \{\mathbf{x}^{n,p} = (x_1^{n,p}, \dots, x_d^{n,p}), p = 1, \dots, P\} \subset \mathbb{R}^d. \quad (3.1)$$

The points of the sets  $X^n$  are computed by employing the Halton's low-discrepancy sequence in  $\mathbb{R}^d$  and standard algorithms for simulating the underlying values in the multidimensional Black-Scholes model.

Now, suppose we want to compute the continuation value of an Bermudan option but only for  $\mathbf{S}_{t_n} = \mathbf{x}^{n,p} \in X^n$ . This goal can be achieved by means of a one step Monte Carlo simulation. In particular, for each  $\mathbf{x}^{n,p} \in X^n$ , we simulate a set of  $M$  points

$$\tilde{X}_p^n = \{\tilde{\mathbf{x}}^{n,p,m} = (\tilde{x}_1^{n,p,m}, \dots, \tilde{x}_d^{n,p,m}), m = 1, \dots, M\} \subset \mathbb{R}^d,$$

which are possible values for  $\mathbf{S}_{t_{n+1}}$  according to the law of  $\mathbf{S}_{t_{n+1}} | \mathbf{S}_{t_n} = \mathbf{x}^{n,p}$ . In particular, for  $i = 1, \dots, d$ ,  $n = 1, \dots, N$ ,  $p = 1, \dots, P$ ,  $m = 1, \dots, M$ , we define

$$\tilde{x}_i^{n,p,m} = x_i^{n,p} e^{(r - \eta_i - \frac{1}{2}\sigma_i^2)\Delta t + \sqrt{\Delta t}\sigma_i \Sigma_i \mathbf{G}^{n,p,m}}, \quad (3.2)$$

where  $\mathbf{G}^{n,p,m} \sim \mathcal{N}(0, I_d)$  is a standard Gaussian random vector and  $\Sigma_i$  is the  $i$ -th row of the matrix  $\Sigma$ , which is defined as a square root of the correlation matrix  $\Gamma$  of the multidimensional Brownian increments. Thus, the risk-less option value can be approximated for each  $\mathbf{x}^{n,p} \in X^n$  by the following scheme:

$$\begin{cases} V_n^{MC}(\mathbf{x}^{n,p}) = \max\left(\frac{e^{-r\Delta t}}{M} \sum_{m=1}^M V_{n+1}^{MC}(\tilde{\mathbf{x}}^{n,p,m}), H(\mathbf{x}^{n,p})\right) & \text{if } n < N, \\ V^{MC}(t_n, \mathbf{x}^{n,p}) = H(\mathbf{x}^{n,p}) & \text{if } n = N. \end{cases} \quad (3.3)$$

Furthermore, the risky value  $\hat{V}_n^{MC}(\mathbf{x}^{n,p})$  for  $M = V$ ,  $t = t_n$  and  $\mathbf{S}_{t_n} = \mathbf{x}^{n,p}$  is approximated by the following scheme:

$$\begin{cases} \hat{V}_n^{MC}(\mathbf{x}^{n,p}) = \max\left\{\hat{C}_n^{MC}(\mathbf{x}^{n,p}), H(\mathbf{x}^{n,p})\right\} & \text{if } n < N, \\ \hat{V}_n^{MC}(\mathbf{x}^{n,p}) = H(\mathbf{x}^{n,p}) & \text{if } n = N, \end{cases} \quad (3.4)$$

with

$$\begin{aligned} \hat{C}_n^{MC}(\mathbf{x}^{n,p}) = \frac{e^{-r_0\Delta t}}{M} \sum_{m=1}^M \left[ \frac{\Delta t}{2} \left( V_{n+1}^{MC}(\tilde{\mathbf{x}}^{n,p,m})^+ c_p + V_{n+1}^{MC}(\tilde{\mathbf{x}}^{n,p,m})^- c_m \right) + \hat{V}_{n+1}^{MC}(\tilde{\mathbf{x}}^{n,p,m}) \right] + \\ + \frac{\Delta t}{2} \left( V_n^{MC}(\mathbf{x}^{n,p})^+ c_p + V_n^{MC}(\mathbf{x}^{n,p})^- c_m \right). \end{aligned}$$

Finally, the risky value  $\hat{V}_n^{MC}(\mathbf{x}^{n,p})$ , for  $M = \hat{V}$ ,  $t = t_n$  and  $\mathbf{S}_{t_n} = \mathbf{x}^{n,p}$ , is computed according to Proposition 2.1, with  $E(\mathbf{x}^{n,p})$  approximated by

$$E^{MC}(\mathbf{x}^{n,p}) = \frac{e^{-r_0\Delta t}}{M} \sum_{m=1}^M \left[ \frac{\Delta t}{2} \left( \hat{V}_{n+1}^{MC}(\tilde{\mathbf{x}}^{n,p,m})^+ c_p + \hat{V}_{n+1}^{MC}(\tilde{\mathbf{x}}^{n,p,m})^- c_m \right) + \hat{V}_{n+1}^{MC}(\tilde{\mathbf{x}}^{n,p,m}) \right]. \quad (3.5)$$

If we proceed backward, the functions  $V_N^{MC}$  and  $\hat{V}_N^{MC}$  are known since they are equal to the payoff of the option  $H$ , so one can compute both  $V_{N-1}^{MC}$  and  $\hat{V}_{N-1}^{MC}$  at  $\tilde{X}_p^n$  by exploiting equations (3.3), (3.4) or (3.5).



Similarly, such a computation at a time step  $t_n$  with  $n < N$  requires the knowledge of the value functions  $V_{n+1}^{MC}$  and  $\hat{V}_{n+1}^{MC}$  at the next time step  $t_{n+1}$  at all the points of the set  $\bigcup_{p=1,\dots,P} \tilde{X}_p^{n+1}$ , but, following the procedure just described, those functions are known only at the points of the set  $X^{n+1}$ : a multidimensional extrapolation tool is required to extending the value functions from  $X^n$  to a suitable neighbourhood of such a set. For this purpose, we exploit Gaussian Process Regression, a class of non-parametric kernel-based probabilistic models that represents the input data as the random observations of a Gaussian stochastic process and it employs a Bayesian approach to perform estimation of the process at new input data. This Machine Learning techniques is well suited to our problem, as it is capable of handling randomly scattered input data and, generally, only a few input observations are needed to obtain accurate predictions. For a brief introduction to GPR, we refer the interested reader to De Spiegeleer et al. [13] or to Goudenège et al. [18], while for a more in-depth discussion, we suggest Rasmussen and Williams [25].

Let  $V_n^{GPR-MC}$  and  $\hat{V}_n^{GPR-MC}$  be the GPR approximations of the functions  $V_n^{MC}$  and  $\hat{V}_n^{MC}$ , obtained from the observations  $\{(\mathbf{x}^{n,p}, V_n^{MC}(\mathbf{x}^{n,p})), p = 1, \dots, P\}$  and  $\{(\mathbf{x}^{n,p}, \hat{V}_n^{MC}(\mathbf{x}^{n,p})), p = 1, \dots, P\}$  respectively. The GPR-MC algorithm requires the replacement of  $V_{n+1}^{MC}$  and  $\hat{V}_{n+1}^{MC}$  in the right side on (3.3), (3.4) or (3.5) with  $V_{n+1}^{GPR-MC}$  and  $\hat{V}_{n+1}^{GPR-MC}$  respectively.

### 3.2 GPR-EI

The GPR-Exact Integration method is similar to the GPR-MC method but the continuation value is estimated through an exact computation of the expectation, based on the Gaussian distribution. By contrast with the GPR-MC method, the predictors employed in the GPR step are related to the logarithms of the underlyings. Secondly, the continuation value at these points is computed through a closed formula which comes from an exact integration.

Here, for the sake of brevity, we limit ourselves to pointing out the main elements of this algorithm, and we refer the interested readers to [18]. The computation of the continuation value, for both risky or riskless options, is a particular case of the computation of an expectation as

$$\mathbb{E}^{\mathbb{Q}} [\Psi(\mathbf{S}_{t+\tau}) | \mathbf{S}_t = \mathbf{x}],$$

with  $\Psi$  a certain function,  $t, t + \tau \in [0, T]$  and  $\tau > 0$ .

Let us define the input set

$$Z = \{\mathbf{z}^p, p = 1, \dots, P\}$$

consisting of  $P$  points in  $\mathbb{R}^d$  quasi-randomly distributed according to the law of the vector  $(\sigma_1 W_\tau^1, \dots, \sigma_d W_\tau^d)^\top$ . In particular, we define

$$\mathbf{z}_i^p = \sqrt{\tau} \sigma_i \Sigma_i \mathbf{h}^p, \quad (3.6)$$

where  $\Sigma_i$  is  $i$ -th row of the matrix  $\Sigma$  and  $\mathbf{h}^p$  is the  $q$ -th point of the Halton's low-discrepancy sequence in  $\mathbb{R}^d$ . Let  $u : Z \rightarrow \mathbb{R}$  be the function defined by

$$u(\mathbf{z}) := \Psi \left( \mathbf{x} \exp \left( \left( r - \boldsymbol{\eta} - \frac{1}{2} \boldsymbol{\sigma}^2 \right) \tau + \mathbf{z} \right) \right). \quad (3.7)$$

The first step is to approximate the function  $u$  by training the GPR method with a Squared Exponential kernel on the set  $Z$ , so that the GPR approximation of the function  $u$  is given by

$$u^{GPR}(\mathbf{z}) = \sum_{p=1}^P k_{SE}(\mathbf{z}^q, \mathbf{z}) \omega_p, \quad (3.8)$$

where  $\omega_1, \dots, \omega_P$  are weights. The continuation value can be computed by integrating the function  $u^{GPR}$  against a  $d$ -dimensional probability density. The use of the Squared Exponential kernel allows one to easily perform such a calculation by means of a closed formula, that is:

$$\mathbb{E}^{\mathbb{Q}} [\Psi(\mathbf{S}_{t+\tau}) | \mathbf{S}_t = \mathbf{x}] \approx \sum_{p=1}^P \omega_p \sigma_f^2 \sigma_l^d \frac{e^{-\frac{1}{2}(\mathbf{z}^p)^\top (\tau \cdot \Pi + \sigma_l^2 I_d)^{-1} (\mathbf{z}^p)}}{\sqrt{\det(\tau \cdot \Pi + \sigma_l^2 I_d)}}, \quad (3.9)$$

where  $\sigma_f$ ,  $\sigma_l$ , and  $\omega_1, \dots, \omega_Q$  are certain constants determined by the GPR approximation of the function  $\mathbf{z} \mapsto u(\mathbf{z})$  considering  $Z$  as the predictor set, and  $\Pi = (\Pi_{i,j})$  is the  $d \times d$  covariance matrix of the vector  $(\sigma_1 W_T^1, \dots, \sigma_d W_T^d)^\top$ , that is  $\Pi_{i,j} = \rho_{i,j} \sigma_i \sigma_j$ .

### 3.3 Control Variates

As suggested by Goudenège et al. [19], control variates technique is a useful tool to improve the accuracy of pricing methods based on GRP. Specifically, we use the European risk-less price  $V^{EU}$  as the control variate for the American risk-less price, and the American risk-less price for the American risky price. In particular, we compute the European risk-less price by Monte Carlo simulations with antithetic variates. We explain the use of control variates technique for the computation of the risk-less American option price  $V$ , and we leave the appropriate adjustments for the risky price  $\hat{V}$  to the reader.

Let  $V^{EU}$  represent the risk-less price of the European option. For a fixed time  $t$  and an underlying stocks value  $\mathbf{x}$ , the American-European price gap is defined as the difference between the American and the European price, that is:

$$v(t, \mathbf{x}) = V(t, \mathbf{x}, 0, 0) - V^{EU}(t, \mathbf{x}, 0, 0). \quad (3.10)$$

The price gap is equal to zero at maturity and, at a general time  $t$ , it can be computed as

$$v(t, \mathbf{x}) = \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}^{\mathbb{Q}} \left[ e^{-r(\tau-t)} K(\tau, \mathbf{S}_\tau) | \mathbf{S}_t = \mathbf{x} \right], \quad (3.11)$$

where  $\mathcal{T}_{t,T}$  stands for the set of all stopping times taking values in  $[t, T]$  and  $K$  is the exercise value gap, defined by

$$K(t, \mathbf{x}) = H(\mathbf{x}) - V^{EU}(t, \mathbf{x}, 0, 0). \quad (3.12)$$

Therefore, the function  $v(t, \mathbf{x})$  can be estimated by exploiting a dynamic programming principle based on Bermudan approximation. In particular, one can use GPR-MC and GPR-EI, by replacing  $H$  with  $K$ . Finally, after computing the initial price gap  $v(0, \mathbf{S}_0)$ , by inverting relation (3.10), one can obtain the American price as

$$V(0, \mathbf{S}_0, 0, 0) = v(0, \mathbf{S}_0) + V^{EU}(0, \mathbf{S}_0, 0, 0). \quad (3.13)$$

*Remark 2.* The computation of the European prices for the control variates technique and the expectation (2.4) are the most time demanding steps. However, these steps can easily be parallelised, thus reducing the total computational time.

## 4 Numerical experiments

In this Section we propose the results of some numerical experiments. The algorithms have been implemented in MATLAB and computations have been performed on a server which employs a 2.40 GHz Intel® Xeon® processor (Gold 6148, Skylake) and 64 GB of RAM. In the remainder of this Section, we discuss 3 American derivatives: a Geometric Put, a Call on the maximum and a Swaption with floor. Table 1 lists all the

Symbol	Meaning	Value	Symbol	Meaning	Value
$S_0^i$	initial spot value	100	$T$	maturity	1.0
$r$	risk free i.r.	0.03	$\lambda_B = \lambda_C$	default intensities	0.04
$\eta_i$	dividend rate	0.00	$R_B = R_C$	recovery rates	0.3
$\sigma_i$	volatility	0.25	$sF$	funding cost	0.028
$\rho_{i,j}$	correlation	0.2	$K$	strike price	100

Table 1: Parameters employed for the numerical experiments in the multi-dimensional Black-Scholes model. In particular,  $s_F = (1 - R_B) \lambda_B$ .

parameters of the stochastic model, with the exception of the dimension  $d$ , which takes on different values from  $d = 2$  up to  $d = 80$ . Based on the results discussed in this Section, one can observe that the two proposed methods are very accurate in the various cases considered. The quality of the results degrades slightly as the size of the problem increases, but the quality of the results is still acceptable, successfully limiting the effects of the curse of dimensionality. Overall, the results proposed by the two methods are always in agreement and very close to the benchmark (when available).

Finally, we stress that obtaining accurate values (in terms of relative error) for the XVA is not an easy task. The XVA is in fact obtained as the difference between two prices that are usually very close to each other. A small estimation error on prices can have a significant weight in relative terms on their difference.

## Geometric Put

We start by considering a Geometric Put option, whose payoff is

$$H(\mathbf{S}_T) = \left( K - \left( \prod_{i=1}^d S_T^i \right)^{\frac{1}{d}} \right)_+.$$

This is a very particularly interesting case since the value of this  $d$ -dimensional option is equal to the value of an appropriate one-dimensional American Put option in the Black-Scholes model, as pointed out in [18, 19]. So, by using one-dimensional standard techniques, such as the CRR tree or a finite difference algorithm, one can obtain very accurate prices for both risk and risk-less American options. In particular, we compute the American benchmark by using both the CRR model with 4000 time steps and a PDE approach with 4000 time steps and 4000 space steps. The obtained values with these two algorithms are equal to three decimal places, so that they can be assumed reliable. The Bermudan benchmark is computed as the American one, but the option has only 41 possible exercise dates, that is  $t_0 = 0, t_1 = 1/40, \dots, t_{40} = 1$ . The GPR-MC method employs 40 time steps, 2000 points and  $10^4$  Monte Carlo simulations, while the GPR-EI method employs 40 time steps and 2000 points.

Table 2 shows the numerical results, which appear to be very accurate and reliable. As far as the price calculation is considered, the relative errors compared to the American benchmark never exceed (in absolute value) 0.14%, which is a very small value. The results are even more interesting when compared to the Bermudan benchmark: in this case, the relative error is always below 0.07%. We can therefore say that, in general, the Bermudan approximation and the algorithmic approximations have a similar contribution to the total error with respect to the American price. The relative error with respect to the XVA are generally larger because the XVA is obtained as the difference of almost equal quantities, so the absolute error must be related to a smaller quantity. However, for the cases considered, the absolute error on the XVA never exceeds

$d$	Option prices			XVA	
	Risk-free	With default risk		$M = V$	$M = \hat{V}$
		$M = V$	$M = \hat{V}$		
American benchmark					
2	6.901	6.659	6.657	0.242	0.244
10	4.866	4.689	4.688	0.177	0.178
20	4.530	4.364	4.363	0.166	0.167
40	4.350	4.190	4.189	0.160	0.161
80	4.257	4.100	4.099	0.157	0.158
Bermudan benchmark					
2	6.895 (-0.09%)	6.651 (-0.12%)	6.649 (-0.12%)	0.244 (0.83%)	0.246 (0.82%)
10	4.863 (-0.06%)	4.685 (-0.09%)	4.683 (-0.11%)	0.178 (0.56%)	0.180 (1.12%)
20	4.527 (-0.07%)	4.360 (-0.09%)	4.358 (-0.11%)	0.167 (0.60%)	0.169 (1.20%)
40	4.347 (-0.07%)	4.186 (-0.10%)	4.185 (-0.10%)	0.161 (0.63%)	0.163 (1.24%)
80	4.254 (-0.07%)	4.096 (-0.10%)	4.095 (-0.10%)	0.158 (0.64%)	0.159 (0.63%)
GPR-MC					
2	6.894 (-0.10%)	6.650 (-0.14%)	6.648 (-0.14%)	0.244 (0.83%)	0.246 (0.82%)
10	4.864 (-0.04%)	4.685 (-0.09%)	4.684 (-0.09%)	0.179 (1.13%)	0.181 (1.69%)
20	4.530 (0.00%)	4.360 (-0.09%)	4.359 (-0.09%)	0.169 (1.81%)	0.171 (2.40%)
40	4.350 (0.00%)	4.187 (-0.07%)	4.185 (-0.10%)	0.164 (2.50%)	0.165 (2.48%)
80	4.255 (-0.07%)	4.097 (-0.07%)	4.095 (-0.10%)	0.159 (1.26%)	0.160 (1.27%)
GPR-EI					
2	6.895 (-0.09%)	6.651 (-0.12%)	6.649 (-0.12%)	0.244 (0.83%)	0.246 (0.82%)
10	4.864 (-0.04%)	4.685 (-0.09%)	4.684 (-0.09%)	0.179 (1.13%)	0.180 (1.12%)
20	4.530 (0.00%)	4.362 (-0.05%)	4.360 (-0.07%)	0.168 (1.20%)	0.170 (1.80%)
40	4.349 (-0.02%)	4.186 (-0.10%)	4.184 (-0.12%)	0.164 (2.50%)	0.165 (2.48%)
80	4.254 (-0.07%)	4.100 (0.00%)	4.097 (0.05%)	0.154 (-1.91%)	0.156 (-1.26%)

Table 2: Numerical results for a Geometric American put option. Values in brackets are the relative errors with respect to the American benchmark.  $d$  stands for the dimension.

Benchmarks			GPR-MC					GPR-EI				
$d$	American	Bermudian	$P$									
			125	250	500	1000	2000	125	250	500	1000	2000
XVA, case $M = V$												
2	0.242	0.244	0.247 (117)	0.245 (202)	0.244 (480)	0.245 (1419)	0.244 (4732)	0.242 (103)	0.244 (126)	0.243 (151)	0.245 (320)	0.244 (841)
10	0.177	0.178	0.183 (160)	0.180 (287)	0.180 (688)	0.179 (1784)	0.179 (5621)	0.180 (132)	0.179 (199)	0.179 (358)	0.179 (698)	0.179 (1803)
20	0.166	0.167	0.171 (243)	0.171 (447)	0.172 (1052)	0.170 (2345)	0.169 (6592)	0.169 (221)	0.169 (344)	0.169 (692)	0.168 (1257)	0.168 (2679)
40	0.160	0.161	0.152 (363)	0.159 (614)	0.164 (1413)	0.165 (3360)	0.164 (8537)	0.159 (360)	0.163 (515)	0.164 (1076)	0.164 (2171)	0.164 (4284)
80	0.157	0.158	0.118 (477)	0.138 (1047)	0.155 (2287)	0.158 (4709)	0.159 (12258)	0.070 (441)	0.119 (1010)	0.145 (1885)	0.151 (3961)	0.154 (7242)
XVA, case $M = \hat{V}$												
2	0.244	0.246	0.249 (125)	0.247 (194)	0.249 (518)	0.248 (1472)	0.246 (4869)	0.244 (97)	0.246 (236)	0.245 (301)	0.247 (607)	0.246 (1570)
10	0.178	0.180	0.188 (190)	0.183 (301)	0.181 (702)	0.182 (1811)	0.181 (5528)	0.182 (132)	0.181 (367)	0.181 (680)	0.181 (1286)	0.181 (3252)
20	0.167	0.169	0.175 (249)	0.175 (464)	0.174 (1011)	0.172 (2295)	0.171 (6680)	0.170 (212)	0.170 (600)	0.170 (1276)	0.170 (2275)	0.170 (4830)
40	0.161	0.163	0.152 (353)	0.163 (657)	0.165 (1345)	0.167 (3365)	0.165 (8782)	0.160 (355)	0.164 (954)	0.166 (1982)	0.165 (3997)	0.165 (7853)
80	0.158	0.159	0.119 (503)	0.141 (1160)	0.156 (2242)	0.159 (5020)	0.160 (12265)	0.071 (832)	0.120 (1804)	0.147 (3366)	0.153 (6826)	0.156 (7395)

Table 3: Numerical results for a Geometric American put option. Values in brackets are the computational times (in seconds).  $P$  is the number of points employed in the GPR algorithms.

2.50% and, in general, tends to increase as the problem size increases. Again, the Bermudian approximation contributes about half of the total error. To conclude, we observe that the results for  $M = V$  and  $M = \hat{V}$  are very similar, both in terms of prices and XVA.

To investigate the convergence rate of the two methods, we compute the XVA by changing the number  $P$  of points employed for the sparse quasi-random grid. As one may observe from the results reported in Table 3, the GPR algorithms provide convergence to Bermudian prices with great accuracy. Moreover, due to the use of the control variate technique, very few points are needed to obtain very accurate results. Obviously, the larger the dimension, the more points are required to approach the exact value. This fact is particularly important as the computational time increases more than linearly as the number of points increases (the higher cost is due to the training of the GPR model, which is cubic). Finally, we note that the GPR-EI method is generally faster and more accurate than GPR-MC, but the latter returns more accurate results in very high dimensions, especially for  $d = 80$ .

## Call on the maximum

The American option Call on the maximum is a difficult to evaluate derivative and so it has been considered by many authors, such as Schoenmakers [27], Lelong [21], Becker et al. [5], Goudenège et al. [18, 19], and Ech-Chafiq et al. [14]. Specifically, the payoff of such an option is given by

$$H(\mathbf{S}_T) = \left( \max_{i=1, \dots, d} S_T^i - K \right)_+.$$

We start the numerical analysis by considering the same model parameters as for the Geometric put, which are reported in Table 1. We stress out that, since the considered derivative is a call option and the

$d$	GPR-MC					GPR-EI					UB
	$P$										
	125	250	500	1000	2000	125	250	500	1000	2000	
XVA, case $M = V$											
2	1.004	1.001	0.998	0.999	0.998	0.999	0.999	0.999	0.998	0.998	$1.009 \pm 0.001$
10	2.191	2.210	2.226	2.231	2.229	2.237	2.231	2.236	2.234	2.230	$2.252 \pm 0.001$
20	2.670	2.720	2.753	2.762	2.771	2.789	2.792	2.788	2.774	2.774	$2.803 \pm 0.001$
40	2.743	3.149	3.223	3.268	3.294	3.317	3.262	3.247	3.296	3.299	$3.337 \pm 0.001$
80	2.674	3.001	3.602	3.702	3.764	2.830	3.241	3.642	3.736	3.741	$3.852 \pm 0.001$
XVA, case $M = \hat{V}$											
2	1.012	1.019	1.011	1.010	1.011	1.010	1.011	1.011	1.010	1.011	$1.021 \pm 0.001$
10	2.217	2.238	2.250	2.254	2.255	2.263	2.257	2.261	2.259	2.257	$2.279 \pm 0.001$
20	2.711	2.748	2.787	2.790	2.803	2.820	2.825	2.823	2.806	2.809	$2.837 \pm 0.001$
40	2.766	3.180	3.259	3.305	3.333	3.358	3.298	3.284	3.336	3.339	$3.377 \pm 0.001$
80	2.705	3.046	3.646	3.742	3.812	2.891	3.259	3.699	3.732	3.783	$3.898 \pm 0.001$

Table 4: Numerical results for a Call on the maximum option. The confidence interval for the upper-bound UB is computed at a 99% confidence level.

underlying pays no dividends ( $\eta_i = 0$ , see Table 1), early exercise is never optimal for the riskless option. Moreover, since the payoff of the derivative is always possible, we can use the closed formulas proposed by Burgard and Kjaer [9] to compute the XVA for the European derivative. Specifically, if  $M = V$ , then

$$XVA^{EU} = V^{EU}(t, \mathbf{S}_0) \cdot \left( 1 - e^{-(\lambda_B + \lambda_C)T} - c_p \frac{1 - e^{-(\lambda_B + \lambda_C)T}}{\lambda_B + \lambda_C} \right),$$

and if  $M = \hat{V}$ , then

$$XVA^{EU} = V^{EU}(t, \mathbf{S}_0) \cdot \left( 1 - e^{(c_p - \lambda_B - \lambda_C)T} \right).$$

It is worth noting that despite the prices of an European and an American riskless options are equals, this does not also apply to their XVAs. In fact, an American option may be exercised early so to reduce the losses due to counterparty default, thus the XVA on the American option is expected to be smaller than the European one. So, we present the XVA on the European option as an upper-bound (UB). Results are shown in Table 4. We can see that both proposed methods provide very accurate values for the cases considered. When a large number of points is used (at least 500), the relative deviation between the returned values, that is the difference divided by the larger value, is less than 2%. The values obtained for XVA are all below the upper-bound, although very close to it.

Finally, for the sake of comparison, let us calculate the XVA for a Call on the maximum considering a positive dividend rate, equal for all underlyings and equal to  $\eta_i = 0.02$ . In this specific case, there are neither benchmarks nor upper-bounds. Table 5 presents the results. In the case under consideration (positive dividend), the valuation seems to be more challenging than in the previous case (with zero dividend). In fact, at least 500 points are needed to obtain a relative deviation of less than 5%.

## Swaption with floor on two portfolios

The derivatives considered in the numerical examples above are all options and therefore their payoff function and their values are always positive. The model considered in this work also admits negative values for the

$d$	GPR-MC					GPR-EI				
	$P$									
	125	250	500	1000	2000	125	250	500	1000	2000
XVA, case $M = V$										
2	0.817	0.805	0.800	0.806	0.804	0.805	0.801	0.802	0.802	0.803
10	1.988	2.003	2.008	2.009	2.006	1.962	1.966	2.000	2.004	1.999
20	2.479	2.529	2.551	2.555	2.556	2.473	2.485	2.481	2.456	2.457
40	2.599	2.948	3.009	3.063	3.080	3.026	2.973	2.964	3.005	2.997
80	2.577	2.850	3.407	3.487	3.557	2.811	3.033	3.461	3.422	3.488
XVA, case $M = \hat{V}$										
2	0.820	0.814	0.809	0.808	0.815	0.814	0.811	0.812	0.811	0.812
10	2.018	2.031	2.027	2.030	2.031	1.981	1.992	2.025	2.026	2.022
20	2.505	2.555	2.581	2.581	2.588	2.503	2.514	2.508	2.488	2.484
40	2.625	2.978	3.051	3.100	3.118	3.059	3.009	2.998	3.040	3.033
80	2.613	2.878	3.449	3.531	3.597	2.848	3.068	3.461	3.464	3.560

Table 5: Numerical results for a Call on the maximum option with a positive dividend rate.

payout, so it is interesting to consider a case with this attribute. Let us now consider an American two-portfolio Swaption with a negative floor, i.e. a derivative in which two portfolios are swapped between counterparties, whose value can be either positive or negative. Specifically, the first portfolio consists of the first  $d/2$  underlyings and the second portfolio consists of the remaining underlyings. For simplicity, we will assume  $d$  to be an even number. In both cases, the underlyings all have the same weight, so the value of each portfolio is equal to the average of the prices of the individual risky assets. The payout of such a derivative is given by

$$H(\mathbf{S}_T) = \max \left( \frac{2}{d} \left( \sum_{i=1}^{d/2} S_T^i - \sum_{i=d/2+1}^d S_T^i \right), K \right).$$

In particular, the floor  $K$  is a negative number, thus the payout of the option can be negative. Table 6 presents the numerical results. We observe that, in the case considered, the estimated values for the XVA are much smaller than in the previous cases. The two methods return very similar values for  $d \leq 40$ , whereas for  $d = 80$  GPR-EI estimates of the XVA are greater than those returned by GPR-MC (approximately +20%). The lack of a benchmark makes it unclear which of the two methods is the more accurate in this case.

## 5 Conclusion

In this paper, we have discussed the problem of calculating the XVA of a derivative that depends on multiple underlyings. This issue plays an essential role in counterparty risk management, also in light of the regulations currently in force. Nevertheless, it is an element that is often overlooked due to the curse of dimensionality associated with the problem of valuing high-dimensional options. Our proposal to address this challenge is to reformulate the problem in probabilistic terms and make use of the GPR-MC and GPR-EI techniques with control variate, which have already been successfully applied in similar contexts. Numerical results show that it is possible to obtain very accurate estimates of the XVA and, in some cases, very few points are

$d$	GPR-MC					GPR-EI				
	$P$									
	125	250	500	1000	2000	125	250	500	1000	2000
XVA, case $M = V$										
2	41.911	42.031	41.953	41.884	42.054	42.086	41.958	42.063	41.982	42.000
10	14.604	14.938	14.693	14.762	14.750	14.820	14.307	14.551	14.691	14.739
20	7.959	8.168	8.567	8.631	8.827	7.612	7.560	8.241	8.230	8.441
40	4.426	4.546	4.277	4.397	4.427	3.710	4.010	4.275	4.215	4.354
80	3.010	1.448	1.855	2.079	2.089	2.173	2.287	2.613	2.524	2.474
XVA, case $M = \hat{V}$										
2	42.560	42.404	42.372	42.563	42.463	42.594	42.377	42.477	42.377	42.549
10	14.894	15.040	14.789	14.849	14.897	14.941	14.369	14.694	14.775	14.854
20	8.022	8.233	8.631	8.653	8.937	7.587	7.726	8.355	8.383	8.592
40	4.446	4.589	4.282	4.414	4.559	3.681	4.081	4.265	4.249	4.412
80	3.179	2.493	1.899	2.104	2.132	2.146	2.301	2.682	2.611	2.593

Table 6: Numerical results for a Swaption with floor on two portfolios. All the results must be multiplied by  $10^{-2}$ .

sufficient to achieve very accurate results. For the considered cases, the proposed methods demonstrate to be effective for large dimensions, thus providing new methods for estimating XVA by overcoming the curse of dimensionality.

## References

- [1] F. Antonelli, A. Ramponi, and S. Scarlatti. Approximate value adjustments for European claims. *European Journal of Operational Research*, 300(3):1149–1161, 2022.
- [2] I. Arregui, B. Salvador, and C. Vázquez. PDE models and numerical methods for total value adjustment in European and American options with counterparty risk. *Applied Mathematics and Computation*, 308:31–53, 2017.
- [3] I. Arregui, B. Salvador, and C. Vázquez. A Monte Carlo approach to American options pricing including counterparty risk. *International Journal of Computer Mathematics*, 96(11):2157–2176, 2019.
- [4] L. Ballotta, G. Fusai, and D. Marazzina. Integrated structural approach to Credit Value Adjustment. *European Journal of Operational Research*, 272(3):1143–1157, 2019.
- [5] S. Becker, P. Cheridito, and A. Jentzen. Deep optimal stopping. *Journal of Machine Learning Research*, 20:74, 2019.
- [6] S. Becker, P. Cheridito, and A. Jentzen. Pricing and hedging American-style options with Deep Learning. *Journal of Risk and Financial Management*, 13(7):158, 2020.
- [7] S. Becker, P. Cheridito, A. Jentzen, and T. Welti. Solving high-dimensional optimal stopping problems using Deep Learning. *European Journal of Applied Mathematics*, 32(3):470–514, 2021.



- [8] D. Brigo and F. Vrina. Disentangling wrong-way risk: pricing credit valuation adjustment via change of measures. *European Journal of Operational Research*, 269(3):1154–1164, 2018.
- [9] C. Burgard and M. Kjaer. PDE representations of options with bilateral counterparty risk and funding costs. *Available at SSRN.com*, 2010.
- [10] S. Crépey and M. F. Dixon. Gaussian process regression for derivative portfolio modeling and application to credit valuation adjustment computations. *Journal of Computational Finance*, 24(1), 2020.
- [11] C. de Graaf, D. Kandhai, and P. Sloot. Efficient estimation of sensitivities for counterparty credit risk with the finite difference Monte Carlo method. *Journal of Computational Finance*, 21(1):83–113, 2017.
- [12] C. S. De Graaf, Q. Feng, D. Kandhai, and C. W. Oosterlee. Efficient computation of exposure profiles for counterparty credit risk. *International Journal of Theoretical and Applied Finance*, 17(04), 2014.
- [13] J. De Spiegeleer, D. B. Madan, S. Reyners, and W. Schoutens. Machine Learning for quantitative finance: fast derivative pricing, hedging and fitting. *Quantitative Finance*, 18(10):1635–1643, 2018.
- [14] Z. E. F. Ech-Chafiq, P. Henry-Labordere, and J. Lelong. Pricing Bermudan options using regression trees/random forests. *arXiv:2201.02587*, 2021.
- [15] Y. Feng. Cva under bates model with stochastic default intensity. *Journal of Mathematical Finance*, 7(3):682–698, 2017.
- [16] A. Gnoatto, A. Picarelli, and C. Reisinger. Deep xVA solver—a neural network based counterparty credit risk management framework. *arXiv:2005.02633*, 2020.
- [17] L. Goudenège, A. Molent, and A. Zanette. Computing credit valuation adjustment solving coupled pides in the bates model. *Computational Management Science*, 17(2):163–178, 2020.
- [18] L. Goudenège, A. Molent, and A. Zanette. Machine learning for pricing American options in high-dimensional Markovian and non-Markovian models. *Quantitative Finance*, 20(4):573–591, 2020.
- [19] L. Goudenège, A. Molent, and A. Zanette. Variance reduction applied to machine learning for pricing Bermudan/American options in high dimension. In O. Kudryavtsev and A. Zanette, editors, *Applications of Lévy Processes*. Nova Science, 2021.
- [20] B. Lapeyre and J. Lelong. Neural network regression for Bermudan option pricing. *Monte Carlo Methods and Applications*, 27(3):227–247, 2021.
- [21] J. Lelong. Dual pricing of american options by Wiener chaos expansion. *SIAM Journal on Financial Mathematics*, 9(2):493–519, 2018.
- [22] M. Ludkovski. Kriging metamodels and experimental design for Bermudan option pricing. *Journal of Computational Finance*, 22(1), 2018.
- [23] V. Piterbarg. Funding beyond discounting: collateral agreements and derivatives pricing. *Risk*, 23(2):97, 2010.
- [24] E. Platen and D. Heath. *A benchmark approach to quantitative finance*. Springer Science & Business Media, 2006.

- [25] C. E. Rasmussen and C. K. Williams. *Gaussian Processes for Machine Learning*, volume 2. MIT Press Cambridge, MA, 2006.
- [26] B. Salvador and C. W. Oosterlee. Total value adjustment for a stochastic volatility model. a comparison with the Black–Scholes model. *Applied Mathematics and Computation*, 391, 2021.
- [27] J. Schoenmakers, J. Zhang, and J. Huang. Optimal dual martingales, their analysis, and application to new algorithms for Bermudan products. *SIAM Journal on Financial Mathematics*, 4(1):86–116, 2013.
- [28] J.-H. She and D. Greco. Neural network for CVA: Learning future values. *arXiv:1811.08726*, 2018.
- [29] G. Yuan, D. Ding, J. Duan, W. Lu, and F. Wu. Total value adjustment of Bermudan option valuation under pure jump Lévy fluctuations. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 32(2), 2022.

## A Proof of Proposition 2.1

Equation (2.7) is a non-linear equation, so, first, we discuss existence and uniqueness of the solution. Let us assume that we have fixed the value of  $t_n$  and  $\mathbf{x}$ , so we can consider them as model parameters. We define the function  $f : \mathbb{R} \rightarrow \mathbb{R}$  as

$$f_{t_n, \mathbf{x}}(z) = \max \left\{ E(\mathbf{x}) + \frac{\Delta t}{2} (z^+ c_p + z^- c_m), H(\mathbf{x}) \right\},$$

so that equation (2.7) can be rewritten as

$$\tilde{V}(t_n, \mathbf{x}) = f_{t_n, \mathbf{x}}(\tilde{V}(t_n, \mathbf{x})).$$

So, in to  $\tilde{V}(t_n, \mathbf{x})$  one has to solve the equation  $z - f_{t_n, \mathbf{x}}(z) = 0$ , that is computing the zeros of the function  $F_{t_n, \mathbf{x}}(z) = z - f_{t_n, \mathbf{x}}(z)$ . We observe that  $F_{t_n, \mathbf{x}}$  is a continuous function and it is piecewise derivable. In particular, if

$$z \neq 0, \quad z \neq 2 \frac{H(\mathbf{x}) - E(\mathbf{x})}{c_p \Delta t}, \quad \text{and} \quad z \neq 2 \frac{H(\mathbf{x}) - E(\mathbf{x})}{c_m \Delta t},$$

then the derivative of  $F_{t_n, \mathbf{x}}$  is given by

$$\frac{d}{dz} (F_{t_n, \mathbf{x}}(z)) = \begin{cases} 1 - \frac{\Delta t}{2} c_p & \text{if } z \geq 0 \text{ and } E(\mathbf{x}) + \frac{\Delta t}{2} c_p z > H(\mathbf{x}), \\ 1 - \frac{\Delta t}{2} c_m & \text{if } z < 0 \text{ and } E(\mathbf{x}) + \frac{\Delta t}{2} c_m z > H(\mathbf{x}), \\ 1 & \text{if } E(\mathbf{x}) + \frac{\Delta t}{2} (z^+ c_p + z^- c_m) < H(\mathbf{x}), \end{cases}$$

that is

$$\frac{d}{dz} (F_{t_n, \mathbf{x}}(z)) = \begin{cases} 1 - \frac{\Delta t}{2} c_p & \text{if } z > 0 \text{ and } z > 2 \frac{H(\mathbf{x}) - E(\mathbf{x})}{\Delta t c_p}, \\ 1 - \frac{\Delta t}{2} c_m & \text{if } z < 0 \text{ and } z > 2 \frac{H(\mathbf{x}) - E(\mathbf{x})}{\Delta t c_m}, \\ 1 & \text{otherwise.} \end{cases}$$

Therefore  $F_{t_n, \mathbf{x}}$  is a continuous piecewise linear function. Moreover, if we assume  $1 - \frac{\Delta t}{2} c_p > 0$  and  $1 - \frac{\Delta t}{2} c_m > 0$  (which is true for  $\Delta t$  small enough)  $F_{t_n, \mathbf{x}}$  is strictly increasing, so it can not have more than one zero. Furthermore, we observe

$$\lim_{z \rightarrow -\infty} F_{t_n, \mathbf{x}}(z) = -\infty, \quad \lim_{z \rightarrow +\infty} F_{t_n, \mathbf{x}}(z) = +\infty,$$

so there is one and only one solution to  $F_{t_n, \mathbf{x}}(z) = 0$ .

Now, we have proved that there is one and only one solution, let us compute it. We rewrite equation (2.7) as

$$\tilde{V}(t_n, \mathbf{x}) = \max \left\{ E(\mathbf{x}) + \frac{\Delta t}{2} \left( \tilde{V}(t_n, \mathbf{x})^+ c_p + \tilde{V}(t_n, \mathbf{x})^- c_m \right), H(\mathbf{x}) \right\}.$$

We distinguish 5 cases.

**Case 1a:**  $\tilde{V}(t_n, \mathbf{x}) = H(\mathbf{x}) \leq 0$ . In this case, we have

$$\max \left\{ E(\mathbf{x}) + \frac{\Delta t}{2} \tilde{V}(t_n, \mathbf{x}), H(\mathbf{x}) \right\} = H(\mathbf{x}),$$

so

$$E(\mathbf{x}) + \frac{\Delta t}{2} H(\mathbf{x}) c_m = E(\mathbf{x}) + \frac{\Delta t}{2} \tilde{V}(t_n, \mathbf{x}) c_m \leq H(\mathbf{x}),$$

thus

$$E(\mathbf{x}) \leq H(\mathbf{x}) \left( 1 - \frac{\Delta t}{2} c_m \right) \leq 0.$$

**Case 1b:**  $H(\mathbf{x}) < \tilde{V}(t_n, \mathbf{x}) \leq 0$ . In this case, we have

$$\max \left\{ E(\mathbf{x}) + \frac{\Delta t}{2} \tilde{V}(t_n, \mathbf{x}) c_m, H(\mathbf{x}) \right\} = E(\mathbf{x}) + \frac{\Delta t}{2} \tilde{V}(t_n, \mathbf{x}) c_m = \tilde{V}(t_n, \mathbf{x}),$$

so

$$\tilde{V}(t_n, \mathbf{x}) = \frac{E(\mathbf{x})}{1 - \frac{\Delta t}{2} c_m},$$

which implies  $E \leq 0$  and

$$H(\mathbf{x}) \left( 1 - \frac{\Delta t}{2} c_m \right) < \tilde{V}(t_n, \mathbf{x}) \left( 1 - \frac{\Delta t}{2} c_m \right) = E(\mathbf{x}) \leq 0.$$

**Case 1c:**  $H(\mathbf{x}) < 0 < \tilde{V}(t_n, \mathbf{x})$ . In this case, we have

$$\max \left\{ E(\mathbf{x}) + \frac{\Delta t}{2} \tilde{V}(t_n, \mathbf{x}) c_p, H(\mathbf{x}) \right\} = E(\mathbf{x}) + \frac{\Delta t}{2} \tilde{V}(t_n, \mathbf{x}) c_p = \tilde{V}(t_n, \mathbf{x}),$$

so

$$\tilde{V}(t_n, \mathbf{x}) = \frac{E(\mathbf{x})}{1 - \frac{\Delta t}{2} c_p},$$

and, since  $E(\mathbf{x}) \geq 0$ , we also have

$$H(\mathbf{x}) \left( 1 - \frac{\Delta t}{2} c_m \right) < 0 < E(\mathbf{x}).$$

**Case 2a:**  $0 < \tilde{V}(t_n, \mathbf{x}) = H(\mathbf{x})$ . In this case, we have

$$\max \left\{ E(\mathbf{x}) + \frac{\Delta t}{2} \tilde{V}(t_n, \mathbf{x}) c_p, H(\mathbf{x}) \right\} = H(\mathbf{x}),$$

so

$$E(\mathbf{x}) + \frac{\Delta t}{2} H(\mathbf{x}) c_p = E(\mathbf{x}) + \frac{\Delta t}{2} \tilde{V}(t_n, \mathbf{x}) c_p \leq H(\mathbf{x}),$$

thus

$$E(\mathbf{x}) \leq H(\mathbf{x}) \left( 1 - \frac{\Delta t}{2} c_p \right).$$

**Case 2b:**  $0 \leq H(\mathbf{x}) < \hat{V}(t_n, S)$ . In this case, we have

$$\max \left\{ E(\mathbf{x}) + \frac{\Delta t}{2} \tilde{V}(t_n, \mathbf{x}) c_p, H(\mathbf{x}) \right\} = E(\mathbf{x}) + \frac{\Delta t}{2} \tilde{V}(t_n, \mathbf{x}) c_p = \tilde{V}(t_n, \mathbf{x}),$$

so

$$\tilde{V}(t_n, \mathbf{x}) = \frac{E(\mathbf{x})}{1 - \frac{\Delta t}{2} c_p},$$

thus

$$E(\mathbf{x}) > H(\mathbf{x}) \left( 1 - \frac{\Delta t}{2} c_p \right).$$

So, cases 1a, 1b, 1c, 2a, 2b, which define a partition of the possible, induce 5 possible relations between  $E(\mathbf{x})$  and  $H(\mathbf{x})$  which are incompatible and exhaustive. Let us summarize these relations:

1. If  $H(\mathbf{x}) \leq 0$  and  $E(\mathbf{x}) \leq H(\mathbf{x}) \left( 1 - \frac{\Delta t}{2} c_m \right) \leq 0$  then case 1a holds and  $\tilde{V}(t_n, \mathbf{x}) = H(\mathbf{x})$ ;
2. If  $H(\mathbf{x}) \leq 0$  and  $H(\mathbf{x}) \left( 1 - \frac{\Delta t}{2} c_m \right) < E(\mathbf{x}) \leq 0$  then case 1b holds and  $\tilde{V}(t_n, \mathbf{x}) = \frac{E(\mathbf{x})}{1 - \frac{\Delta t}{2} c_m}$ ;
3. If  $H(\mathbf{x}) \leq 0$  and  $0 < E(\mathbf{x})$  then case 1c holds and  $\tilde{V}(t_n, \mathbf{x}) = \frac{E(\mathbf{x})}{1 - \frac{\Delta t}{2} c_p}$ ;
4. If  $H(\mathbf{x}) > 0$  and  $E(\mathbf{x}) \leq H(\mathbf{x}) \left( 1 - \frac{\Delta t}{2} c_p \right)$  then case 2a holds and  $\tilde{V}(t_n, \mathbf{x}) = H(\mathbf{x})$ ;
5. If  $H(\mathbf{x}) > 0$  and  $E(\mathbf{x}) > H(\mathbf{x}) \left( 1 - \frac{\Delta t}{2} c_p \right)$  then case 2b holds and  $\tilde{V}(t_n, \mathbf{x}) = \frac{E(\mathbf{x})}{1 - \frac{\Delta t}{2} c_p}$ .

These 5 cases solve the fixed point problem (2.5).