

# PRICING HIGH-DIMENSIONAL BERMUDAN OPTIONS USING VARIANCE-REDUCED MONTE CARLO METHODS

Peter Hepperger \*

We present a numerical method for pricing Bermudan options depending on a large number of underlyings. The asset prices are modeled with exponential time-inhomogeneous jump-diffusion processes. We improve the least-squares Monte Carlo method proposed by Longstaff and Schwartz introducing an efficient variance reduction scheme. A control variable is obtained from a low-dimensional approximation of the multivariate Bermudan option. To this end, we adapt a model reduction method called proper orthogonal decomposition (POD), which is closely related to principal component analysis, to the case of Bermudan options. Our goal is to make use of the correlation structure of the assets in an optimal way. We compute the expectation of the control variable by either solving a low-dimensional partial integro-differential equation or by applying Fourier methods. The POD approximation can also be used as a candidate for the minimizing martingale in the dual pricing approach suggested by Rogers. We evaluate both approaches in numerical experiments.

KEY WORDS Bermudan options, dimension reduction, proper orthogonal decomposition, regression-based Monte Carlo, Fourier methods

## 1 INTRODUCTION

The present article is concerned with numerical pricing of multidimensional Bermudan options. We call an option multidimensional if it is written on more than one underlying. Important examples on the stock market include index options on an average, and basket options on the minimum or maximum of a set of assets. Derivatives on other markets depending on several prices or driving factors also fall into this category. We will consider Bermudan options with a finite number of exercise dates. By choosing this number sufficiently high, the method presented here can be used

---

\*Zentrum Mathematik, Technische Universität München, 85748 Garching bei München, Germany (hepperger@ma.tum.de). Supported by the International Graduate School of Science and Engineering (IGSSE) of Technische Universität München.

to approximate American options with a continuum of exercise possibilities. Note that in fact whenever Monte Carlo (MC) simulation is used for pricing American options, this actually amounts to an approximation with a Bermudan option on the MC time discretization grid. The approximation error may increase, however, for a larger number of exercise points.

We assume the underlying prices to be driven by a multivariate time-inhomogeneous jump-diffusion process. Theoretically, the same numerical pricing methods as in the one-dimensional case can be employed here. These include partial integro-differential equations (PIDEs) [7, 8] and Fourier transform methods [4, 5, 18]. However, they both suffer from the *curse of dimensionality* which means that in practice they cannot be applied directly when the number of underlyings is large. MC methods are a feasible alternative, whose complexity does not increase exponentially in the dimension. MC simulations for Bermudan options are often based on dynamic programming principles. The Snell envelope is obtained through backward recursion. The conditional expectation of future cashflows at each exercise point can be approximated using regression, a method first introduced by Longstaff and Schwartz [19, 3, 15]. The option is exercised if and only if the current intrinsic value is larger than this expectation. These methods yield an approximation from below for the fair price. A different approach uses a dual representation of the price, which allows for the computation of an upper bound [20]. For an overview of Bermudan MC pricing see [16] and the references therein.

The major drawback of all MC algorithms is their comparatively slow convergence rate. There are several techniques, subsumed under the term *variance reduction*, which can help obtaining more accurate results with fewer simulated paths. These include antithetic variables, importance sampling, and control variables [1, 10, 11]. We will focus on the latter and present an improved Longstaff-Schwartz algorithm using a low-dimensional approximation of the option price as the control variable.

The dimension reduction relies on an orthogonal projection method called proper orthogonal decomposition (POD) [17]. Similar to principal component analysis (PCA), a small set of orthonormal vectors is found, which minimizes the projection error. The approximation is hence optimal (in the  $L^2$ -sense). A similar projection method has already been successfully applied to European option pricing [13]. We extend the concept to Bermudan options. In particular, we no longer rely on direct, accurate computation of the option price via POD, but rather use a fast, coarse approximation for variance reduction purposes.

An estimate for the approximation error is derived. The expectation of the low-dimensional POD approximation can be computed efficiently with PIDE or Fourier algorithms. Once this is done, the solution can be used for two purposes: first, it can serve as a control variable. As the approximation is highly correlated with the full-dimensional price process, this results in a substantially decreased variance of the modified MC estimator. Second, the POD solution is a candidate for the minimizing martingale in Rogers' [20] dual MC method. We discuss both approaches. The dimension of the projection can be freely chosen. This allows for a trade-off between reduced variance and increased effort for the computation of the low-dimensional expectation.

The effectiveness of the POD depends on the correlation structure of the underlying assets. High correlation allows for a more efficient decrease of variance with fewer POD components. We investigate the performance of the POD variance-reduction in numerical experiments. Using different types of options, basket sizes, correlation parameters, and numbers of exercise dates, we show that the overall computational time can as a rule be reduced by at least 50%, often even by more than 80%.

The paper is organized as follows. In Section 2, we present the multivariate jump-diffusion model and formulate the Bermudan option pricing problem. The dimension reduction methods and the corresponding convergence results are derived in Section 3. Next, we describe the algorithms for variance-reduced Bermudan MC and the dual MC approach in detail in Section 4. Section 5 contains the numerical experiments. We describe the test settings and analyze the computational results. Finally, Section 6 gives a short conclusion and summary of the article.

## 2 BERMUDAN BASKET OPTIONS

In this section, the market model used throughout the paper is introduced. We define the driving stochastic jump-diffusion process, declare assumptions concerning the coefficients, and state the Bermudan option pricing problem.

### 2.1 THE ASSET PRICE PROCESS

We consider a Bermudan option depending on  $n$  assets. The terminal date of maturity is  $T > 0$  (last exercise date). The multivariate asset price process is denoted

$$(2.1) \quad S_t := (S_1(t), \dots, S_n(t)) := (S_1(0) e^{X_1(t)}, \dots, S_n(0) e^{X_n(t)}) \in \mathbb{R}^n, \quad t \in [0, T].$$

For each  $i = 1, \dots, n$ , the price  $S_i$  of the  $i$ th asset is modeled as the product of its initial value  $S_i(0) > 0$  and the ordinary exponential of a time-inhomogeneous jump-diffusion process  $X_i$ , given by

$$(2.2) \quad \begin{aligned} X_t &:= (X_1(t), \dots, X_n(t)) \\ &:= \int_0^t \gamma_s ds + \int_0^t \sigma_s dW(s) + \int_0^t \int_H \eta_s \xi \tilde{M}(d\xi, ds) \in \mathbb{R}^n, \quad t \in [0, T]. \end{aligned}$$

The diffusion part is driven by an  $\mathbb{R}^n$ -valued Brownian motion  $W$ . The jumps are characterized by  $\tilde{M}$ , the compensated random measure of an  $\mathbb{R}^n$ -valued compound Poisson process

$$J_t = \sum_{i=1}^{N_t} Y_i, \quad t \geq 0,$$

which is independent of  $W$ . Here,  $N$  denotes a Poisson process with intensity  $\lambda$  and  $Y_i \sim P^Y$  ( $i = 1, 2, \dots$ ) are iid on  $\mathbb{R}^n$  (and independent of  $N$ ). The corresponding Lévy

measure is denoted by  $\nu = \lambda P^Y$ . We assume the drift  $\gamma : [0, T] \rightarrow \mathbb{R}^n$ , the volatility  $\sigma : [0, T] \rightarrow \mathbb{R}^{n \times n}$ , and the jump integrand  $\eta : [0, T] \rightarrow \mathbb{R}^{n \times n}$  to be deterministic functions. We make the following assumption concerning the moments of the process.

**Assumption 2.1.** *The second exponential moment of the jump distribution  $Y$  exists:*

$$E[e^{2\|Y\|_{\mathbb{R}^n}}] = \int_{\mathbb{R}^n} e^{2\|\xi\|_{\mathbb{R}^n}} P^Y(d\xi) < \infty.$$

We assume further that

$$\int_0^T \|\gamma_t\|_{\mathbb{R}^n}^2 dt < \infty, \quad \int_0^T \|\sigma_t\|_{\mathbb{R}^{n \times n}}^2 dt < \infty, \quad \text{and} \quad \|\eta_t\|_{\mathbb{R}^{n \times n}} \leq 1 \quad \text{for a.e. } t \in [0, T].$$

The interest rate  $r$  is assumed to be constant. In order to avoid a discussion of possible measure changes, we suppose the model (2.1) to be stated under the pricing measure. In view of the dimension reduction performed in Section 3, it turns out to be useful to express the value of the option in terms of the centered process

$$Z_t := X_t - E[X_t], \quad t \in [0, T].$$

Since the asset price process  $S = S(Z)$  depends on  $Z$  in a deterministic way, this is nothing more than a simple transform of variables.

## 2.2 BERMUDAN OPTIONS

A Bermudan option grants the holder the right to exercise at one of  $N_{ex} \in \mathbb{N}$  admissible dates, which we denote by  $0 \leq t_1 < t_2 < \dots < t_{N_{ex}} = T$ . Let  $\mathcal{T}(t, T)$  denote the set of all stopping times with values in  $\{t_i | 1 \leq i \leq N_{ex} \text{ and } t_i \geq t\}$ . For simplicity, we assume a constant interest rate  $r > 0$ . The discounted value  $V$  of a Bermudan option at time  $t$ , given that the option was not yet exercised, is the solution of the optimal stopping problem

$$(2.3) \quad V(t, z) = \sup_{\tau \in \mathcal{T}(t, T)} E[e^{-r\tau} g(S(Z_\tau)) | Z_t = z].$$

The payoff  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  is determined by the asset price process  $S$  which can be expressed in terms of the centered jump-diffusion  $Z$ . We make the following assumption.

**Assumption 2.2.** *We assume that there is a constant  $L_g$  such that the payoff function  $g$  satisfies the Lipschitz condition*

$$|g(S) - g(\tilde{S})| \leq L_g \|S - \tilde{S}\|_{\mathbb{R}^n} \quad \text{for every } S, \tilde{S} \in \mathbb{R}^n.$$

**Remark 2.3.** *Assumption 2.2 is considerably weaker than the corresponding assumption in [13, Ass. 3.4]: it refers to the payoff in terms of  $S$  instead of  $Z$ . In particular, it is satisfied for plain vanilla call and put options on weighted averages of the asset prices. The convergence proofs below (Lemma 3.4 and Theorem 3.5) account for this weaker assumption with additional technical estimates.*

This assumption is, in particular, satisfied for index options written on a weighted average of the assets. An index put has the intrinsic value

$$g(S) = \left( K - \sum_{i=1}^n w_i S_i \right)^+,$$

where  $K$  is the strike and  $w_i \in \mathbb{R}$  are constant weights. Other examples include maximum or minimum options. A maximum or minimum put corresponds to the payoff

$$g(S) = \left( K - \max_{i=1, \dots, n} S_i \right)^+ \quad \text{or} \quad g(S) = \left( K - \min_{i=1, \dots, n} S_i \right)^+,$$

respectively.

The aim when pricing Bermudan options is to find the optimal exercise time for (2.3). It is well known that this can be done by backward dynamic programming: at time  $t = T$  the value of the option is

$$(2.4) \quad V(T, z) = e^{-rT} g(S(Z_T)).$$

For any previous exercise date  $t_i$ ,  $i = 0, \dots, N_{ex} - 1$ , the value is

$$(2.5) \quad V(t_i, z) = \max \left\{ e^{-rt_i} g(S(z)), E [V(t_{i+1}, Z_{t_{i+1}}) | Z_{t_i} = z] \right\}.$$

Hence, it is optimal to exercise at time  $t_i$  if and only if the intrinsic value  $g(S(z))$  is larger than or equal to the expected discounted future cash flow (given the option is not yet exercised). Computing the conditional expectations

$$(2.6) \quad E [V(t_{i+1}, Z_{t_{i+1}}) | Z_{t_i} = z]$$

for every exercise date is the basic challenge. The fair value of the option at time  $t = 0$  is then given by  $V(0, 0)$ .

### 3 DIMENSION REDUCTION

It is possible to derive a partial integro-differential (PIDE) equation which is satisfied by the conditional expectation (2.6) (see, e.g., [13]). Such a differential equation, however, suffers from the curse of dimensionality. The same holds true for Fourier transforms. It is hardly possible to apply these methods directly for large values of  $n$ , say  $n > 10$ . In this section, we derive a low-dimensional approximation for Bermudan options. To this end, we employ proper orthogonal decomposition (POD), which makes use of the correlation of the individual assets. The idea is similar to principal component analysis (PCA): we approximate the centered driving process  $Z$  with a small set of orthonormal vectors.

**Definition 3.1.** A sequence of orthonormal vectors  $\{p_l\}_{l=1,\dots,n} \subset \mathbb{R}^n$  is called a POD-basis for  $Z_T$ , if it solves the minimization problem

$$\min_{\langle p_i, p_j \rangle_{\mathbb{R}^n} = \delta_{ij}} E \left\| Z_T - \sum_{l=1}^d p_l \langle Z_T, p_l \rangle_{\mathbb{R}^n} \right\|_{\mathbb{R}^n}^2$$

for every  $d = 1, \dots, n$ .

The number  $d$  of components in the projection will later be the dimension of the approximating problem. Like in PCA we can obtain a POD-basis by solving an eigenvalue problem. Let  $C_{X_T}$  be the covariance matrix of  $X_T$ , and thus also of  $Z_T$ . The following proposition shows that the eigenvectors of  $C_{X_T}$  are indeed a POD-basis. It is quoted from [13, Thm. 3.3].

**Proposition 3.2.** Every sequence of orthonormal eigenvectors  $(p_l)_{l=1,\dots,n}$  of the covariance  $C_{X_T}$ , ordered by descending size of the corresponding eigenvalues  $\mu_1 \geq \mu_2 \geq \dots \geq 0$ , is a POD-basis. The expectation of the projection error is

$$(3.1) \quad E \left\| Z_T - \sum_{l=1}^d p_l \langle Z_T, p_l \rangle_{\mathbb{R}^n} \right\|_{\mathbb{R}^n}^2 = \sum_{l=d+1}^n \mu_l.$$

Subsequently, let  $(p_l)_{l \in \mathbb{N}}$  and  $(\mu_l)_{l=1,\dots,n}$  denote the orthonormal basis and eigenvalues from Proposition 3.2. We define the projection operator

$$\mathcal{P}_d : \begin{cases} \mathbb{R}^n & \rightarrow \text{span}\{p_1, p_2, \dots, p_d\} \cong \mathbb{R}^d, \\ z & \mapsto \sum_{l=1}^d \langle z, p_l \rangle_{\mathbb{R}^n} p_l. \end{cases}$$

Hence, we can rewrite (3.1) as

$$E \left\| Z_T - \mathcal{P}_d Z_T \right\|_{\mathbb{R}^n}^2 = \sum_{l=d+1}^n \mu_l.$$

So far, we have approximated the value of  $Z$  only at time  $T$ . It turns out, however, that this is indeed sufficient to obtain small projection errors for arbitrary  $t \in [0, T]$ .

**Proposition 3.3.** The following holds:

$$(3.2) \quad \sup_{t \in [0, T]} E \left\| Z_t - \mathcal{P}_d Z_t \right\|_{\mathbb{R}^n}^2 \leq \sum_{l=d+1}^n \mu_l.$$

*Proof.* This is a direct consequence of the independent increments of  $Z$ . Using the Pythagorean theorem, we obtain

$$\begin{aligned} E \left\| Z_T - \mathcal{P}_d Z_T \right\|_{\mathbb{R}^n}^2 &= E \left\| Z_t - \mathcal{P}_d Z_t + (Z_T - Z_t) - \mathcal{P}_d (Z_T - Z_t) \right\|_{\mathbb{R}^n}^2 \\ &= E \left\| Z_t - \mathcal{P}_d Z_t \right\|_{\mathbb{R}^n}^2 + E \left\| (Z_T - Z_t) - \mathcal{P}_d (Z_T - Z_t) \right\|_{\mathbb{R}^n}^2 \\ &\geq E \left\| Z_t - \mathcal{P}_d Z_t \right\|_{\mathbb{R}^n}^2. \end{aligned}$$

Applying Proposition 3.2 yields (3.2).  $\square$

Consequently, it is not necessary to change Definition 3.1 in order to approximate every  $Z_t$ ,  $t \in [0, T]$ . This is due to the fact that by approximating  $Z_T$ , we capture also the events up to time  $T$ . We use the projection to define the function

$$(3.3) \quad V_d(t, z_d) = \sup_{\tau \in \mathcal{T}(t, T)} E[e^{-r\tau} g(S(\mathcal{P}_d Z_\tau)) | \mathcal{P}_d Z_t = z_d]$$

for every  $z_d \in \mathbb{R}^d$ ,  $t \in [0, T]$ . This is in fact the price process of a  $d$ -dimensional Bermudan option. If  $d$  is chosen sufficiently small, this price can be computed efficiently with Fourier or PIDE methods. In our numerical experiments (compare Section 5),  $d \leq 3$  gave good results. Since the payoff in (3.3) is different from the payoff in (2.3), the corresponding optimal stopping times need not be identical. Nevertheless, we can state an error estimate for the difference  $|V_d(0, 0) - V(0, 0)|$ . In order to show this convergence result, we will make use of the following lemma. It is concerned with the approximation error for a fixed stopping time.

**Lemma 3.4.** *Let  $\tau \in \mathcal{T}(0, T)$  be a fixed stopping time and let  $\mu_1 \geq \mu_2 \geq \dots \geq 0$  be the eigenvalues of the covariance matrix  $C_{X_\tau}$ . Then there exists a constant  $C > 0$  (independent of  $d$ ) such that*

$$\left| E[e^{-r\tau} g(S(\mathcal{P}_d Z_\tau))] - E[e^{-r\tau} g(S(Z_\tau))] \right| \leq C \sqrt{\sum_{l=d+1}^n \mu_l}$$

for every  $d = 1, \dots, n$ .

*Proof.* Throughout the proof, we denote every constant factor depending on  $T$  but not on  $d$  by  $C$ , i.e., the value of  $C$  is not fixed. Using Assumption 2.2 and the definition of  $S$ , we obtain

$$\begin{aligned} & \left| E[e^{-r\tau} g(S(\mathcal{P}_d Z_\tau))] - E[e^{-r\tau} g(S(Z_\tau))] \right| \\ & \leq E \left[ L_g \|S(\mathcal{P}_d Z_\tau) - S(Z_\tau)\|_{\mathbb{R}^n} \right] \\ & = L_g E \left\| \left( S_i(0) e^{\int_0^\tau \gamma_t(i) dt} \left| e^{(\mathcal{P}_d Z_\tau)(i)} - e^{Z_\tau(i)} \right| \right)_{i=1}^n \right\|_{\mathbb{R}^n}. \end{aligned}$$

Since all norms on  $\mathbb{R}^n$  are equivalent, we can use the 1-norm and obtain

$$(3.4) \quad \begin{aligned} & \left| E[e^{-r\tau} g(S(\mathcal{P}_d Z_\tau))] - E[e^{-r\tau} g(S(Z_\tau))] \right| \\ & \leq C E \left[ \sum_{i=1}^n S_i(0) e^{\int_0^\tau \gamma_t(i) dt} \left| e^{(\mathcal{P}_d Z_\tau)(i)} - e^{Z_\tau(i)} \right| \right]. \end{aligned}$$

For the term depending on  $\gamma$ , we use Assumption 2.1 and obtain

$$\left| \int_0^\tau \gamma_t(i) dt \right| \leq \int_0^T \|\gamma_t\|_{\mathbb{R}^n} dt \leq C \left( \int_0^T \|\gamma_t\|_{\mathbb{R}^n}^2 dt \right)^{\frac{1}{2}} \leq C.$$

Next, we apply the mean-value theorem to the exponential function for the estimate

$$\left| e^{(\mathcal{P}_d Z_\tau)(i)} - e^{Z_\tau(i)} \right| \leq e^{\max\{(\mathcal{P}_d Z_\tau)(i), Z_\tau(i)\}} \|\mathcal{P}_d Z_\tau - Z_\tau\|_{\mathbb{R}^n}.$$

Inserting these estimates into (3.4) and applying the Cauchy–Schwarz inequality yields

$$\begin{aligned} & \left| E[e^{-r\tau} g(S(\mathcal{P}_d Z_\tau))] - E[e^{-r\tau} g(S(Z_\tau))] \right| \\ (3.5) \quad & \leq C \sum_{i=1}^n S_i(0) E \left[ e^{\max\{(\mathcal{P}_d Z_\tau)(i), Z_\tau(i)\}} \|\mathcal{P}_d Z_\tau - Z_\tau\|_{\mathbb{R}^n} \right] \\ & \leq C \sum_{i=1}^n S_i(0) \left( E \left[ e^{2 \max\{(\mathcal{P}_d Z_\tau)(i), Z_\tau(i)\}} \right] \right)^{\frac{1}{2}} \left( E \|\mathcal{P}_d Z_\tau - Z_\tau\|_{\mathbb{R}^n}^2 \right)^{\frac{1}{2}}. \end{aligned}$$

In order to get rid of the stopping time  $\tau$ , we will make use of Doob’s inequality. The centered process  $Z$  is a martingale by construction and so is  $\mathcal{P}_d Z$ , since  $\mathcal{P}_d$  is a linear operator. The norm function is convex. Hence,  $\|\mathcal{P}_d Z_\tau - Z_\tau\|_{\mathbb{R}^n}$  is a non-negative submartingale. Doob’s inequality yields

$$(3.6) \quad E \|\mathcal{P}_d Z_\tau - Z_\tau\|_{\mathbb{R}^n}^2 \leq E \left( \sup_{t \in [0, T]} \|\mathcal{P}_d Z_t - Z_t\|_{\mathbb{R}^n} \right)^2 \leq 4E \|\mathcal{P}_d Z_T - Z_T\|_{\mathbb{R}^n}^2.$$

For the exponential term, we find

$$e^{\max\{(\mathcal{P}_d Z_\tau)(i), Z_\tau(i)\}} \leq e^{(\mathcal{P}_d Z_\tau)(i)} + e^{Z_\tau(i)} = e^{\langle Z_\tau, \mathcal{P}_d e_i \rangle_{\mathbb{R}^n}} + e^{\langle Z_\tau, e_i \rangle_{\mathbb{R}^n}},$$

where  $e_i$  denotes the  $i$ th standard unit vector. Since the exponential function is convex, both  $(e^{\langle Z_t, \mathcal{P}_d e_i \rangle_{\mathbb{R}^n}})_{t \in [0, T]}$  and  $(e^{\langle Z_t, e_i \rangle_{\mathbb{R}^n}})_{t \in [0, T]}$  are submartingales. Using Young’s and Doob’s inequalities, we get

$$\begin{aligned} E \left[ e^{2 \max\{(\mathcal{P}_d Z_\tau)(i), Z_\tau(i)\}} \right] & \leq E \left[ \left( e^{\langle Z_\tau, \mathcal{P}_d e_i \rangle_{\mathbb{R}^n}} + e^{\langle Z_\tau, e_i \rangle_{\mathbb{R}^n}} \right)^2 \right] \\ & \leq 2E \left[ \left( e^{\langle Z_\tau, \mathcal{P}_d e_i \rangle_{\mathbb{R}^n}} \right)^2 + \left( e^{\langle Z_\tau, e_i \rangle_{\mathbb{R}^n}} \right)^2 \right] \\ & \leq 8E \left[ e^{2 \langle Z_T, \mathcal{P}_d e_i \rangle_{\mathbb{R}^n}} + e^{2 \langle Z_T, e_i \rangle_{\mathbb{R}^n}} \right] \end{aligned}$$

With [14, Prop. 2.3], we obtain

$$(3.7) \quad E \left[ e^{2 \max\{(\mathcal{P}_d Z_\tau)(i), Z_\tau(i)\}} \right] \leq C.$$

Inserting (3.6) and (3.7) in (3.5) yields

$$\left| E[e^{-r\tau} g(S(\mathcal{P}_d Z_\tau))] - E[e^{-r\tau} g(S(Z_\tau))] \right| \leq C \left( E \|\mathcal{P}_d Z_T - Z_T\|_{\mathbb{R}^n}^2 \right)^{\frac{1}{2}} \sum_{i=1}^n S_i(0).$$

Including the sum  $\sum_{i=1}^n S_i(0)$  in the constant  $C$  and applying Proposition 3.2 to the remaining expectation concludes the proof.  $\square$



The following theorem states the main convergence result for  $V_d$ .

**Theorem 3.5.** *Let  $\mu_1 \geq \mu_2 \geq \dots \geq 0$  be the eigenvalues of the covariance matrix  $C_{X_T}$ . Then there exists a constant  $C > 0$  such that*

$$|V_d(0,0) - V(0,0)| \leq C \sqrt{\sum_{l=d+1}^n \mu_l}$$

for every  $d = 1, \dots, n$ .

*Proof.* Let

$$\tau_n := \operatorname{argsup}_{\tau \in \mathcal{T}(t,T)} E[e^{-r\tau} g(S(Z_\tau))]$$

and

$$\tau_d := \operatorname{argsup}_{\tau \in \mathcal{T}(t,T)} E[e^{-r\tau} g(S(\mathcal{P}_d Z_\tau))]$$

be optimal stopping times for  $V(0,0)$  and  $V_d(0,0)$ , respectively. Then we have

$$(3.8) \quad E[e^{-r\tau_n} g(S(Z_{\tau_n}))] \geq E[e^{-r\tau_d} g(S(Z_{\tau_d}))]$$

and

$$(3.9) \quad E[e^{-r\tau_d} g(S(\mathcal{P}_d Z_{\tau_d}))] \geq E[e^{-r\tau_n} g(S(\mathcal{P}_d Z_{\tau_n}))]$$

by construction. Moreover, we know from Lemma 3.4 that

$$(3.10) \quad \left| E[e^{-r\tau_x} g(S(\mathcal{P}_d Z_{\tau_x}))] - E[e^{-r\tau_x} g(S(Z_{\tau_x}))] \right| \leq C \sqrt{\sum_{l=d+1}^n \mu_l}$$

for  $\tau_x \in \{\tau_n, \tau_d\}$ . Combining (3.10) and (3.8), we find

$$\begin{aligned} E[e^{-r\tau_d} g(S(\mathcal{P}_d Z_{\tau_d}))] &\leq E[e^{-r\tau_d} g(S(Z_{\tau_d}))] + C \sqrt{\sum_{l=d+1}^n \mu_l} \\ &\leq E[e^{-r\tau_n} g(S(Z_{\tau_n}))] + C \sqrt{\sum_{l=d+1}^n \mu_l}. \end{aligned}$$

On the other hand, using (3.9) and (3.10), we get

$$\begin{aligned} E[e^{-r\tau_d} g(S(\mathcal{P}_d Z_{\tau_d}))] &\geq E[e^{-r\tau_n} g(S(\mathcal{P}_d Z_{\tau_n}))] \\ &\geq E[e^{-r\tau_n} g(S(Z_{\tau_n}))] - C \sqrt{\sum_{l=d+1}^n \mu_l}. \end{aligned}$$

Together, these estimates yield

$$|V_d(0,0) - V(0,0)| = \left| E[e^{-r\tau_d} g(S(\mathcal{P}_d Z_{\tau_d}))] - E[e^{-r\tau_n} g(S(Z_{\tau_n}))] \right| \leq C \sqrt{\sum_{l=d+1}^n \mu_l}.$$

□

Theorem 3.5 shows that  $V_d(0,0)$  is a good approximation for the true Bermudan option value, if  $d$  is chosen sufficiently large. The convergence estimate does not depend on the number of exercise points. Moreover, it can be shown that the eigenvalues  $\mu_i$ ,  $i = 1, \dots, n$  decay at least like a power function (or even exponentially fast), if certain smoothness criteria for the covariance matrix are satisfied (see [13, Th. 3.14]). If, on the other hand, the individual assets are entirely independent, the POD method will not yield any improvement. The dimension reduction relies on the correlation of the basket.

## 4 IMPROVING LEAST-SQUARES MONTE CARLO

For European options, it is possible to solve the projected optimal stopping problem (3.3) for relatively large dimensions  $d$  using Fourier or PIDE methods on sparse grids [13, Th. 3.14]. Sparse grids, though, can be applied efficiently only for smooth functions. In the European case, the necessary differentiability properties of  $V_d$  are due to the smoothing effect of the diffusive part of the process. For Bermudan options, on the other hand, a non-differentiable maximum function is involved at every exercise point. The resulting solutions are not sufficiently smooth for sparse grid convergence results, and the effect gets worse if the number  $N_{ex}$  of exercise points is increased. Moreover, the condition number of the linear equation systems of the corresponding time discretized PIDEs may increase substantially. Hence, the use of sparse grids in the context of American options poses considerable theoretical and practical challenges.

We will instead rely on conventional full grids to solve (3.3), using dimensions  $d \leq 3$ . According to Theorem 3.5, the accuracy of  $V_d(0,0)$  as an approximation for  $V(0,0)$  is then of course limited. However,  $V_d$  is in any case highly correlated with  $V$ . In this section, we discuss how to exploit this property to improve the convergence rate of MC methods, using either variance reduced least-squares MC or a duality based pricing approach.

### 4.1 VARIANCE REDUCTION

Pricing Bermudan options with MC simulations is more demanding than pricing European contracts. We will use the well established least-squares MC method introduced by Longstaff and Schwartz [19]. Starting with the terminal value (2.4) at time  $T$ , we iterate backwards ( $i = N_{ex} - 1, N_{ex} - 2, \dots, 1$ ) over all previous exercise dates using the recursive formula (2.5). We denote by  $(\mathcal{F}_{t_i})_{i=1}^{N_{ex}}$  the natural filtration corresponding to the driving process  $Z$ . The conditional expectation

$$(4.1) \quad E [V(t_{i+1}, Z_{t_{i+1}}) | \mathcal{F}_{t_i}]$$

is assumed to be a linear combination of a set of  $\mathcal{F}_{t_i}$ -adapted basis variables. This set of variables may, e.g., include the current state  $Z_{t_i}(1), \dots, Z_{t_i}(n)$  of the driving process, the values  $S_{t_i}(1), \dots, S_{t_i}(n)$  of the assets, and the value  $F(Z_{t_i})$  of the function on which

the option is written (the weighted average for index options, or the minimum or maximum for the corresponding put options). Polynomials in all of these quantities are also possible candidates. In practice, it turns out that the method is rather insensitive to the concrete choice of basis variables.

Once a set of variables is fixed, the basis variables are evaluated in each recursion step for each path of the MC simulation. The discounted future cashflow for each path when the option is not exercised is already known from the backward recursion. The conditional expectation (4.1) is then computed from simple linear regression *over all paths*. This is the major difference to European options: we make use of the information from all paths at the same time. Since the linear regression amounts to solving a symmetric linear equation system, the computational effort of the regression step is not linear in the number of paths. Moreover, it is harder to do these computations in parallel. The computational time of the whole least-squares algorithm, however, is often dominated by the simulation part (compare also the numerical experiments in Section 5).

As with all MC methods, a rather large number of paths is needed to obtain accurate approximations. For a discussion of convergence rates, see [2]. We denote the number of simulated paths by  $N$ . Let  $V_j^N, j = 1, \dots, N$ , be the price computed for the  $j$ th simulated path, using the exercise strategy given by least-squares MC. Once the number of paths and the exercise policy are fixed, these are independent and identically distributed (i.i.d.) copies of a random variable  $V^N$ . The MC estimator for the option price  $V(0,0)$  is given by

$$\theta = \frac{1}{N} \sum_{j=1}^N V_j^N.$$

In order to improve the estimate, we will employ variance reduction with a control variable as described, e.g., in [10, chap. 1.3]. This amounts to finding a second random variable  $U^N$ , which is closely related to  $V^N$ , but whose expectation  $E[U^N]$  can be computed much more efficiently. Then, we choose  $\alpha \in \mathbb{R}$  and compute the new estimator

$$\theta_{vr}(\alpha) = \frac{1}{N} \sum_{j=1}^N \left[ V_j^N + \alpha \left( U_j^N - E[U^N] \right) \right],$$

where  $U_j^N$  are i.i.d. copies of  $U^N$ . The expectation of  $\theta_{vr}(\alpha)$  is obviously identical to that of  $\theta$ , therefore no bias is introduced. The variance of the new estimator is given by

$$\text{Var}(\theta_{vr}(\alpha)) = \frac{1}{N} \left( \text{Var}(V^N) + 2\alpha \text{Cov}(V^N, U^N) + \alpha^2 \text{Var}(U^N) \right).$$

The minimal possible variance

(4.2)

$$\theta_{vr}(\alpha^*) = \frac{1}{N} \left( \text{Var}(V^N) - \frac{\text{Cov}^2(V^N, U^N)}{\text{Var}(U^N)} \right) = \frac{1}{N} \text{Var}(V^N) \left( 1 - \text{Corr}^2(V^N, U^N) \right)$$

is obtained for

$$\alpha^* = -\frac{\text{Cov}(V^N, U^N)}{\text{Var}(U^N)}.$$

This optimal value  $\alpha^*$  cannot be calculated directly in practice, since at least the covariance  $\text{Cov}(V^N, U^N)$  is usually unknown. Thus,  $\alpha^*$  has to be estimated. Since we simulate values of  $V^N$  and  $U^N$  anyway, this requires no additional effort. We can use the empirical estimates for  $\text{Cov}(V^N, U^N)$  and  $\text{Var}(U^N)$ . Equation (4.2) shows that the variance of the improved estimator will be arbitrarily small, if the correlation  $\text{Corr}(V^N, U^N)$  of  $V^N$  and  $U^N$  is large.

We will employ the dimension reduction presented in the previous section to obtain a suitable choice for  $U^N$ . The solution  $V_d$  of the projected Bermudan pricing problem (3.3) converges to  $V$  according to Theorem 3.5. Therefore, we set  $U^N$  equal to the approximated value of  $V_d$  along each path. We can obtain this approximation with the very same method we use for  $V$ . Once we have simulated  $N$  paths for  $Z$  and approximated the pathwise value of  $V$  with the least-squares method, obtaining the value of  $V_d$  is possible with small additional effort for two reasons: first, we reuse the same paths and can thus skip the expensive simulation step. Second, since  $V_d$  effectively depends only on the  $d$ -dimensional process  $\mathcal{P}_d Z$ , and  $d \ll n$ , we can reduce the number of basis variables for the regression significantly without losing accuracy.

The main benefit of this choice of the control variable  $U^N$  is that any method suitable for low-dimensional Bermudan options can be used to compute the expectation  $V_d(0, 0)$ . The steps required for this variance-reduced least-squares MC method are summarized in Algorithm 1.

There is a large variety of algorithms which can be used to compute the expectation  $E[U^N] = V_d(0, 0)$  numerically, most notably those based on PIDEs or on Fourier transforms. These approaches are well known; we will not discuss them in detail. An overview of available methods can be found, e.g., in [6, 10]. PIDE methods make use of the fact that the function  $f(z) := E[V_d(t_{i+1}, Z_{t_{i+1}}) | \mathcal{P}_d Z_{t_i} = z]$  satisfies the  $d$ -dimensional PIDE

$$\begin{aligned} -\partial_t f(z) &= \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d (\sigma_t \sigma_t^T)_{ij} \partial_{z_i z_j}^2 f(z) \\ &\quad + \int_{\mathbb{R}^d} \left\{ f(z + \eta_t \zeta) - f(z) - \sum_{i=1}^d (\eta_t \zeta)_i \partial_{z_i} f(z) \right\} \nu(d\zeta), \end{aligned}$$

for  $z \in \mathbb{R}^d$ ,  $t \in [t_k, t_{k+1}]$ , and  $k \in 1, \dots, N_{ex} - 1$ . The differentials can be discretized using finite elements or finite differences. The integral term requires numerical quadrature. Details of the implementation are discussed in [13].

Alternatively, the function  $f$  can be represented in terms of Fourier transforms:

$$f(z) = \frac{e^{-rt_{i+1}}}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-iz^T u} \varphi_{t_i, t_{i+1}}(-u) \int_{\mathbb{R}^d} e^{iu^T y} g(S(y)) dy du$$

---

**Algorithm 1** Variance-reduced least-squares MC pricing

---

```
1: compute expectation  $V_d(0,0) = \sup_{\tau \in \mathcal{T}(0,T)} E[e^{-r\tau} g(S(\mathcal{P}_d Z_\tau))]$  (PIDE or FFT)
2: simulate  $N$  paths of  $Z$ 
3: compute projected paths  $\mathcal{P}_d Z$ 
4: for  $j = 1, 2, \dots, N$  do // terminal values
5:    $V_j^N = e^{-rT} g(S(Z_T(j)))$ 
6:    $U_j^N = e^{-rT} g(S(\mathcal{P}_d Z_T(j)))$ 
7: end for
8: for  $i = N_{ex} - 1, N_{ex} - 2, \dots, 1$  do // backward recursion
9:   regression with large basis set for  $E[V_j^N | Z_{t_i}]$ 
10:  regression with small basis set for  $E[U_j^N | \mathcal{P}_d Z_{t_i}]$ 
11:  for  $j = 1, 2, \dots, N$  do // pathwise exercise decision
12:    compute intrinsic values  $e^{-rt_i} g(S(Z_{t_i}(j)))$  and  $e^{-rt_i} g(S(\mathcal{P}_d Z_{t_i}(j)))$ 
13:     $V_j^N = \max \left\{ e^{-rt_i} g(S(Z(j))), E[V_j^N | Z_{t_i}] \right\}$ 
14:     $U_j^N = \max \left\{ e^{-rt_i} g(S(\mathcal{P}_d Z(j))), E[U_j^N | \mathcal{P}_d Z_{t_i}] \right\}$ 
15:  end for
16: end for
17: estimate mean  $\mu_V$  and variance  $\sigma_V^2$  of  $V^N$ 
18: estimate mean  $\mu_U$  and variance  $\sigma_U^2$  of  $U^N$ 
19:  $\alpha = \frac{1}{N-1} \sum_{j=1}^N \frac{(V_j^N - \mu_V)(U_j^N - \mu_U)}{\sigma_V \sigma_U}$ 
20:  $\theta_{vr}(\alpha) = \frac{1}{N} \sum_{j=1}^N \left[ V_j^N + \alpha (U_j^N - V_d(0,0)) \right]$ 
21: return  $\theta_{vr}(\alpha)$ 
```

---

for  $z \in \mathbb{R}^d$  and  $i \in 1, \dots, N_{ex} - 1$ , where  $\varphi_{t_i, t_{i+1}}$  denotes the characteristic function of the increment of  $Z$  over the interval  $[t_i, t_{i+1}]$ . In general, the payoff  $g$  is not integrable. This can be remedied by either truncating the payoff or by subtracting a suitable smooth function. The Fourier transforms can then be computed very efficiently using fast Fourier transforms (FFT). Option pricing with Fourier methods is discussed, e.g., in [4, 18].

## 4.2 DUAL METHOD

Since the exercise policy found by linear regression of the conditional expectation is not necessarily optimal, the least-squares MC presented in the previous section actually computes an (arbitrarily precise) lower bound for the option price. The same value is of course also a lower bound for the value of the continuously exercisable American option. The approach presented by Rogers [20] uses a duality argument to obtain an

approximation from above. The Bermudan option price can be written as

$$(4.3) \quad V(0,0) = \inf_{M \in \mathcal{M}_0^1} E \left[ \sup_{i=1, \dots, N_{ex}} \left( e^{-rt_i} g(S(Z_{t_i})) - M_{t_i} \right) \right],$$

where  $\mathcal{M}_0^1$  is the space of all martingales  $M$  satisfying  $M_0 = 0$  and

$$E \left[ \sup_{i=1, \dots, N_{ex}} |M_{t_i}| \right] < \infty.$$

Rogers' idea is to pick a suitable martingale  $M$  and compute the expectation in (4.3) via MC simulation. Since the chosen martingale will in general not be optimal, the computed price is larger than the infimum and is thus an upper bound. According to [12, Thm. 1], we could also insert a supermartingale into the dual formulation and still obtain an upper bound.

The choice of a "good" martingale  $M \in \mathcal{M}_0^1$  is a delicate issue, since it is related to the increments of a hedging strategy for the option. In practice, however, any martingale related to the price of the option may yield remarkably accurate results. Since the projected price process is an approximation for the true price, we suggest setting

$$M_t := \begin{cases} V_d(t, \mathcal{P}_d Z_t) - V_d(0,0), & t \in [0, \tau], \\ e^{-r\tau} g(S(\mathcal{P}_d Z_\tau)) - V_d(0,0), & t \in (\tau, T], \end{cases}$$

where  $\tau$  denotes the optimal stopping time for the projected Bermudan option. Note that  $V_d(\tau, \mathcal{P}_d Z_\tau) = e^{-r\tau} g(S(\mathcal{P}_d Z_\tau))$  holds by construction. Moreover,  $M_t + V_d(0,0)$  can be interpreted as the discounted value of a European option which grants the holder the same wealth as an optimally exercised Bermudan option at maturity. Thus, it is a martingale under the pricing measure. Algorithm 2 lists the computational steps of the proposed dual method.

For practical applications, we are of course interested in bounds which are sufficiently sharp to serve as approximations of the true price. Therefore, we compare the dual MC method and the variance-reduced MC method with respect to computational speed and accuracy. In our numerical experiments (see Section 5), the dual method showed extremely fast convergence. It is sufficient to simulate a very small number of paths for  $Z$ . In addition, there is no regression step. The individual paths can be processed completely in parallel. There are some caveats, though. In contrast to the variance reduction method, where we have used only the value  $V_d(0,0)$ , we have to compute the solution  $V_d$  on a full space-time discretization grid in order to evaluate  $M$ . As before, this can be done with PIDE and FFT methods, but computing the full solution has several disadvantages. First, memory consumption is increased considerably. Second, since we have to truncate the computational domain to a compact subset of  $\mathbb{R}^d$ , the accuracy of the solution decreases when we approach the boundary of this

---

**Algorithm 2** Duality based MC pricing

---

```
1: compute the solution  $V_d$  on a space-time grid (PIDE or FFT)
2: simulate  $N$  paths of  $Z$ 
3: compute projected paths  $\mathcal{P}_d Z$ 
4: for  $j = 1, 2, \dots, N$  do // pathwise maximum
5:   initial value  $v(j) = g(S(0))$ 
6:   exercised_projected = false
7:   for  $i = 1, 2, \dots, N_{ex}$  do // forward iteration
8:     intrinsic value  $u = e^{-rt_i} g(S(Z_{t_i}(j)))$ 
9:     if not exercised_projected then
10:      interpolate projected continuation value  $M = V_d(t_i, \mathcal{P}_d Z_{t_i}(j)) - V_d(0, 0)$ 
11:      intrinsic projected value  $u_d = e^{-rt_i} g(S(\mathcal{P}_d Z_{t_i}(j)))$ 
12:      if  $u_d \geq M$  then
13:        exercised_projected = true
14:      end if
15:    end if
16:     $v(j) = \max\{v(j), u - M\}$ 
17:  end for
18: end for
19:  $\theta_{dual} = \frac{1}{N} \sum_{j=1}^N v(j)$ 
20: return  $\theta_{dual}$ 
```

---

subset. Moreover, the evaluation of  $M_t$  at arbitrary arguments requires interpolation and, thus, usually a finer discretization grid.

The most important drawback of the dual method is probably that the precision of the result depends on the chosen martingale  $M$ . If  $M$  is too far from optimal, we cannot expect convergence to the true option value. In particular, we will see that for large numbers of assets with moderate correlation and many exercise possibilities, we would have to choose a large value for  $d$  in order to obtain a reasonable sharp bound. If, on the other hand, correlation is sufficiently high, the dual method is by far superior regarding computational speed.

## 5 NUMERICAL EXPERIMENTS

In this section, we analyze the performance of the dimension reduction approach in numerical experiments. The variance reduced and dual MC methods are applied to test problems with various parameters. We vary the number of assets in the basket and the number of exercise dates. We price options on baskets with high or low correlation (compared to real stock market data) and study two different types of options.

## 5.1 TEST SETTING

We consider baskets of  $n = 10, 20,$  and  $30$  assets. The driving stochastic process for our test problems is a jump-diffusion of the form (2.2) with time-constant volatility and jump distribution. Similar to [21], we include independent jumps for each individual asset as well as common jumps for all assets. The common jumps for all assets are driven by a compound Poisson process with intensity  $\lambda_0$  and are of fixed relative height  $\eta_0$ . The additional individual jumps of each asset price have intensity  $\lambda_i$  and relative height  $\eta_i, i = 1, \dots, n$ . The price process of each asset  $S_i$  satisfies

$$\frac{dS_i(t)}{S_i(t)} = rdt + \sum_{j=1}^n \sigma_{ij} dW_j(t) + \eta_0 d[N_0(t) - \lambda_0 t] + \eta_i d[N_i(t) - \lambda_i t].$$

The Brownian motions  $W_j, j = 1, \dots, n$ , as well as the Poisson processes  $N_0, N_1, \dots, N_n$  are all independent. The entries  $(\sigma_{ij})_{i,j=1}^n$  of the volatility matrix are chosen such that the covariance  $\mathcal{C}_D \in \mathbb{R}^{n \times n}$  of the diffusion part satisfies

$$\mathcal{C}_D(i, j) = (\sigma \sigma^T)_{ij} = 0.2 e^{-\rho|i-j|}, \quad \text{for } i, j = 1, \dots, n,$$

where  $\rho \in \mathbb{R}$  is a parameter controlling the decay of correlation. The discounted value of every asset is a martingale under the pricing measure and can be written as exponential of a jump-diffusion process as follows:

$$S_i(t) = S_i(0) \exp \left\{ \left( r - \frac{1}{2} \sum_{j=1}^n \sigma_{ij}^2 - \eta_0 \lambda_0 - \eta_i \lambda_i \right) t + \sum_{j=1}^n \sigma_{ij} W_j(t) + \ln(1 + \eta_0) N_0(t) + \ln(1 + \eta_i) N_i(t) \right\}, \quad i = 1, \dots, n.$$

We set  $\lambda_0 = \lambda_1 = \dots = \lambda_n = 1, \eta_1 = \dots = \eta_n = -0.05$ . For the remaining correlation and jump parameters we use two different sets of values:

$$\begin{array}{ll} \text{High correlation} & \rho = 0.1 \quad \eta_0 = -0.15, \\ \text{Low correlation} & \rho = 0.4 \quad \eta_0 = -0.10. \end{array}$$

Hence, we have a faster decaying correlation and less pronounced common jumps in the “low correlation” scenario.

Figure 1 shows the eigenvalues of the covariance matrix of  $Z_T$  in both scenarios. All values are divided by the largest eigenvalue  $\mu_1$  for normalization. The decay is exponential. A faster decay means higher correlation and, thus, usually better performance of the dimension reduction method. For comparison, the eigenvalues obtained from the empirical covariance of the top 20 S&P 500 stocks are also plotted. The graph shows that these eigenvalues are between those obtained from the test problem with the two parameter sets described above.



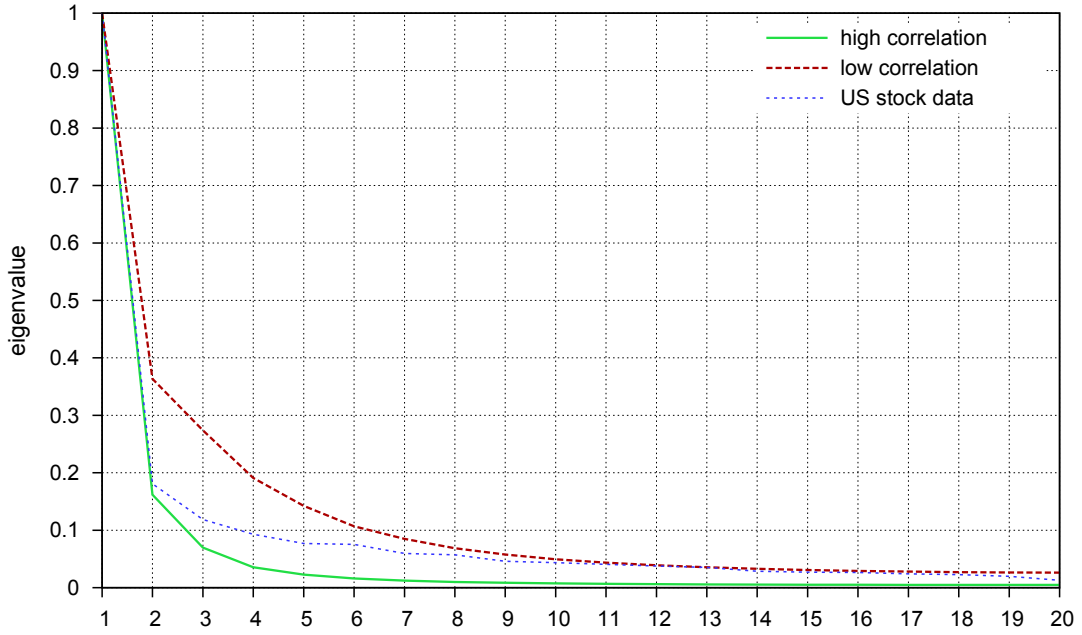


Fig. 1: Eigenvalue decay for different correlation scenarios and S&P 500 stocks. The eigenvalues of the covariance matrices are ordered by size and normalized.

The initial value for the assets is  $S_i(0) = 50$ ,  $i = 1, \dots, n$ . We compute the price of put options with strike  $K = 50$  and value

$$V(0,0) = \sup_{\tau \in \mathcal{T}(0,T)} E[e^{-r\tau} (K - F(S(Z_\tau)))^+],$$

where  $F$  is either the average

$$F_{avg}(S) := \frac{1}{n} \sum_{i=1}^n S_i$$

or the minimum

$$F_{min}(S) := \min_{i=1, \dots, n} S_i$$

The admissible exercise dates are equally spaced:  $t_i = \frac{i}{N_{ex}}$ ,  $i = 1, \dots, N_{ex}$ . Their total number is either  $N_{ex} = 10$  or  $N_{ex} = 100$ .

For the computation of  $V_d$  both PIDE and FFT methods have been tested. Both yield very similar results. Since the FFT showed slightly superior accuracy on identical grids in our test cases, all of the results below refer to the FFT method. The grid refinement and domain truncation are chosen in such a way that the absolute error of  $V_d(0,0)$  is well below 0.005 (0.5 cent). Usually,  $2^6$  grid points in each coordinate are sufficient to achieve this. The complete method was implemented in C++, using the FFTW code [9] for the Fourier transforms. The code was parallelized for shared memory systems with OpenMP and executed on a workstation with 8 Opteron processors at 2.7 GHz.

## 5.2 RESULTS

**COMPUTATIONAL TIME** Before we analyze the gain in precision obtained with variance reduction and dual MC, we examine the additional computational cost per path which is needed for these methods. If we fix the number  $N$  of MC paths, doing a plain least-squares MC is obviously less time consuming than computing additional control variables. This extra work is only worth the effort, if we can reduce the number of paths significantly, so that the total computing time needed to achieve a given precision decreases. On the other hand, the dual method may take less time per path, since no regression is needed and the pathwise computation of maxima is inexpensive. Figure 2 gives an overview of the computational times needed for different parts of the algorithm. The simulation of paths is identical for all tested methods. The regression step for the control variables takes slightly less time than the regression for the original MC data, since it uses less basis variables. The set of basis variables for the full MC contains  $n + 3$  values  $(Z_1, \dots, Z_n, F, F^2, F^3)$ , while the set for the control variable has only  $d + 3$  elements  $((\mathcal{P}_d Z)_1, \dots, (\mathcal{P}_d Z)_d, F, F^2, F^3)$ . The cost of the FFT increases exponentially in the dimension  $d$ . Using full grids with dimensions  $d > 3$  implies computational times which are larger than the original MC method, even after taking the reduced number of paths into account.

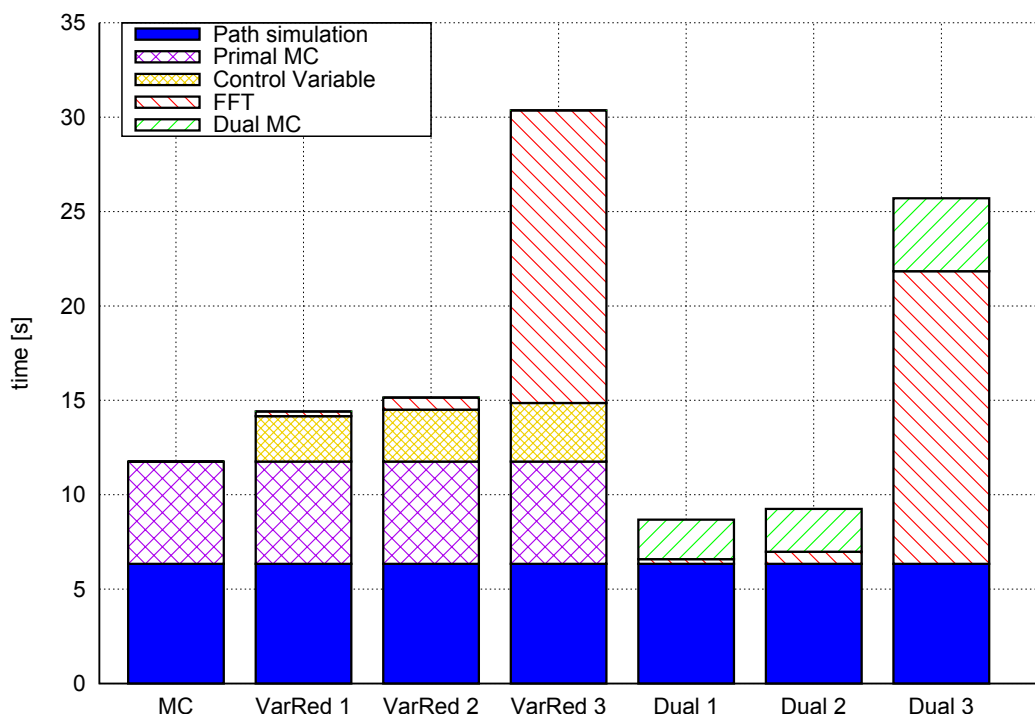


Fig. 2: Computational times using variance reduction (VarRed) or dual MC (Dual) for  $N = 50000$  paths. (Average put option, low correlation,  $n = 30$ ,  $N_{ex} = 100$ .) The number behind the method indicates the projected dimension  $d$ .

**REDUCED VARIANCE** Since we have now seen how much additional time the variance reduction needs, the question is how much accuracy we gain from it. A measure for the accuracy is the variance of the MC prices. It can be used to obtain bounds on the precision, e.g., by Chebyshev’s inequality. If the number of paths is sufficiently large and the computed exercise policies do not change substantially, when further paths are added, the variance is inversely proportional to the number of paths. Consequently, half the variance means that roughly half the number of simulated paths is sufficient for identical precision.

Figures 3 and 4 show results of the variance reduction and the dual MC for different settings. In general, the methods work better for the average option than for the minimum option. This is not surprising, because the average is captured better by the POD components than the minimum. The variance can of course be decreased further if the correlation is high. The number of exercise dates  $N_{ex}$  also has a small influence. More exercise points yield less effective variance reduction. The dual method converges extremely fast in every setting. Its accuracy, however, is only satisfying for highly correlated baskets and the average put option. In any other case, a higher value for  $d$  ( $d > 3$ ) and a finer grid for the FFT solution are needed to obtain a reasonable approximation with the dual method. This is usually not worth the computational effort. The variance reduction, on the other hand, always works, although its effect is hardly visible in the plot for the “worst case” of a minimum put option on lowly correlated assets.

**CHOICE OF DIMENSION** The effectiveness of the variance reduction of course also depends on the dimension  $d$  of the projected problem. For the average option, it turns out that  $d = 1$  already yields a substantial improvement over the plain least-squares MC method without variance reduction. Increasing the dimension gives only slightly smaller variances. Nevertheless,  $d = 2$  is worth considering, since the additional computational effort is small. For the minimum put option, increasing  $d$  has a much larger impact. Adding further POD components decreases the variance significantly. Figure 5 illustrates this effect. In terms of overall computational effort,  $d = 2$  turns out to be a good choice, although  $d = 3$  is sometimes even better (depending on the number of assets and the efficiency of the method used to compute  $V_d(0, 0)$ ).

In order to obtain a good estimate for the variance of the MC price, a large number of MC experiments (usually at least 1000), each with the given number of paths  $N$ , is necessary. A good approximation can be obtained by computing the variance *within* each set of  $N$  paths and dividing by  $N$ . In fact, this approximation is accurate if the exercise policy does not vary for different sets of simulated paths. In order to rule out effects due to a possible change of exercise policy between different sets of paths, we use the mean of this value over 100 MC experiments (with  $N$  paths each).

VARIANCE AND TIME RATIOS Tables 1 and 2 summarize a large number of computational results for average and minimum options, respectively. Each cell of these tables contains two numbers. The first one is the ratio of the variance  $\sigma_{vr}^2$  after variance reduction to the variance  $\sigma_{MC}^2$  of the plain least-squares MC method. The second one is the ratio of total computing time  $t_{vr}$  with variance reduction to computing time  $t_{MC}$  using plain MC for  $N = 100000$  paths. While computational time increases with dimension  $d$ , the variance ratio decreases. The product of the two quantities gives a very rough estimate of the total computing time ratio, because the number of paths needed for a certain accuracy of the result decreases proportional to the variance. Taking, e.g., the entry for  $N_{ex} = 10$ ,  $n = 10$ , and  $d = 2$  in Table 1, we obtain  $0.04 \cdot 1.26 = 0.05$ , which means that we can save about 95% of computing time with variance reduction of dimension 2. The entry for  $N_{ex} = 100$ ,  $n = 30$ , and  $d = 2$  in Table 2, on the other hand, yields  $0.49 \cdot 1.19 = 0.58$ , corresponding to a 42% lower computational cost due to variance reduction.

		$N_{ex} = 10$			$N_{ex} = 100$		
	# assets	$d = 1$	$d = 2$	$d = 3$	$d = 1$	$d = 2$	$d = 3$
high corr	10	0.04 1.26	0.03 1.40	0.03 1.62	0.05 1.22	0.04 1.29	0.04 1.52
	20	0.06 1.24	0.05 1.31	0.03 1.63	0.08 1.21	0.06 1.30	0.05 1.65
	30	0.07 1.19	0.06 1.26	0.05 1.83	0.10 1.19	0.07 1.25	0.07 1.58
low corr	10	0.10 1.26	0.06 1.34	0.04 2.44	0.12 1.22	0.07 1.30	0.05 2.19
	20	0.14 1.22	0.10 1.29	0.08 2.40	0.15 1.21	0.10 1.28	0.08 2.03
	30	0.16 1.19	0.12 1.25	0.10 2.43	0.18 1.21	0.13 1.24	0.11 1.92

Table 1: Variance ratio  $\frac{\sigma_{vr}^2}{\sigma_{MC}^2}$  and time ratio  $\frac{t_{vr}}{t_{MC}}$  for *average put option* ( $N = 100000$  paths).

		$N_{ex} = 10$			$N_{ex} = 100$		
	# assets	$d = 1$	$d = 2$	$d = 3$	$d = 1$	$d = 2$	$d = 3$
high corr	10	0.46 1.22	0.20 1.30	0.13 1.60	0.48 1.16	0.23 1.27	0.15 1.49
	20	0.52 1.19	0.26 1.26	0.14 1.57	0.54 1.17	0.29 1.24	0.20 1.55
	30	0.55 1.15	0.29 1.21	0.19 1.66	0.68 1.14	0.43 1.21	0.32 1.44
low corr	10	0.64 1.20	0.36 1.30	0.24 2.23	0.66 1.18	0.39 1.27	0.26 1.97
	20	0.69 1.17	0.44 1.25	0.32 2.09	0.70 1.16	0.47 1.24	0.35 1.81
	30	0.70 1.14	0.47 1.20	0.37 2.07	0.71 1.13	0.49 1.19	0.39 1.67

Table 2: Variance ratio  $\frac{\sigma_{vr}^2}{\sigma_{MC}^2}$  and time ratio  $\frac{t_{vr}}{t_{MC}}$  for *minimum put option* ( $N = 100000$  paths).

## 6 CONCLUSION

In this article, we have presented a dimension reduction method for high-dimensional Bermudan options under jump-diffusion models. A low-dimensional approximation for the option price is obtained by orthogonal projection to a suitable set of basis vectors, making use of the correlation structure of the assets. An error estimate for the approximation has been shown. The expectation of the approximated price process can be computed with any algorithm suitable for pricing low-dimensional Bermudan options, in particular Fourier or PIDE methods. The solution then is used as a control variable for variance reduction of the Longstaff–Schwartz MC algorithm. It can also serve as a candidate for the dual MC method proposed by Rogers. Numerical experiments show that the dual method is sufficiently accurate only for highly correlated baskets and options with a moderate number of exercise points. In these cases, its convergence rate is outstandingly fast. It is, however, not suitable to approximate American options with a continuum of exercise dates. The variance-reduced MC does not suffer from these restrictions. Its convergence is slower, but it still yields significant improvements in overall computational cost. The stronger the correlation of the underlyings and the higher the dimension of the projected equation, the better the variance reduction works. Like the original least-squares MC simulation, the presented variance-reduced least-squares MC method can be used to approximate American options with continuous exercise possibilities by choosing a sufficiently large number of discrete exercise dates, at the cost of a possibly increasing approximation error.

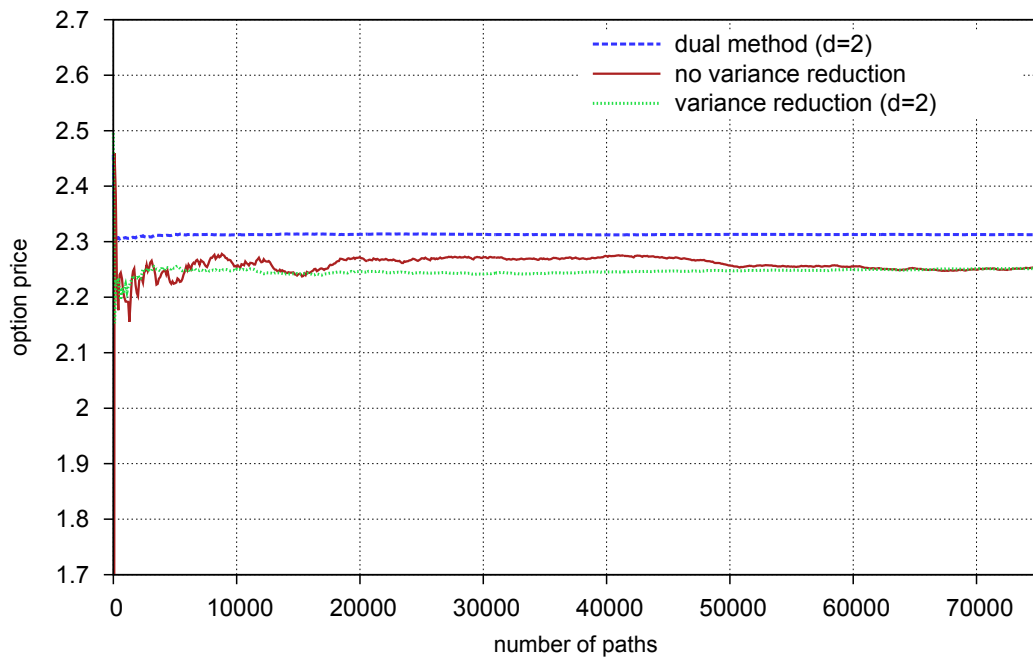
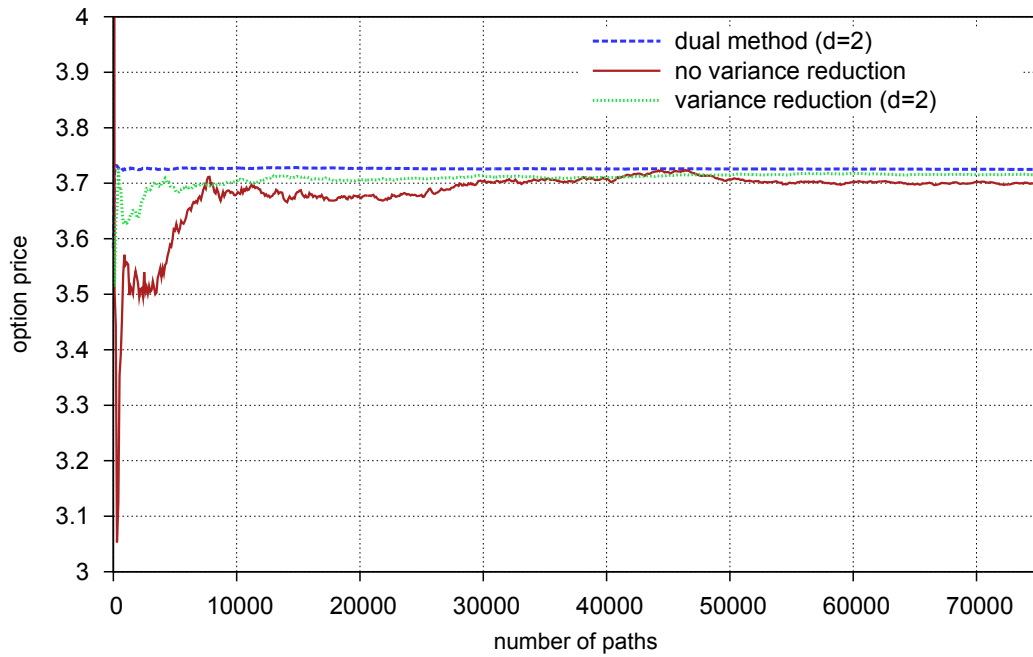


Fig. 3: MC, variance reduced MC, and dual MC for *average put option* ( $n = 30, d = 2$ ).  
 Top: high correlation,  $N_{ex} = 10$ ; bottom: low correlation,  $N_{ex} = 100$ .

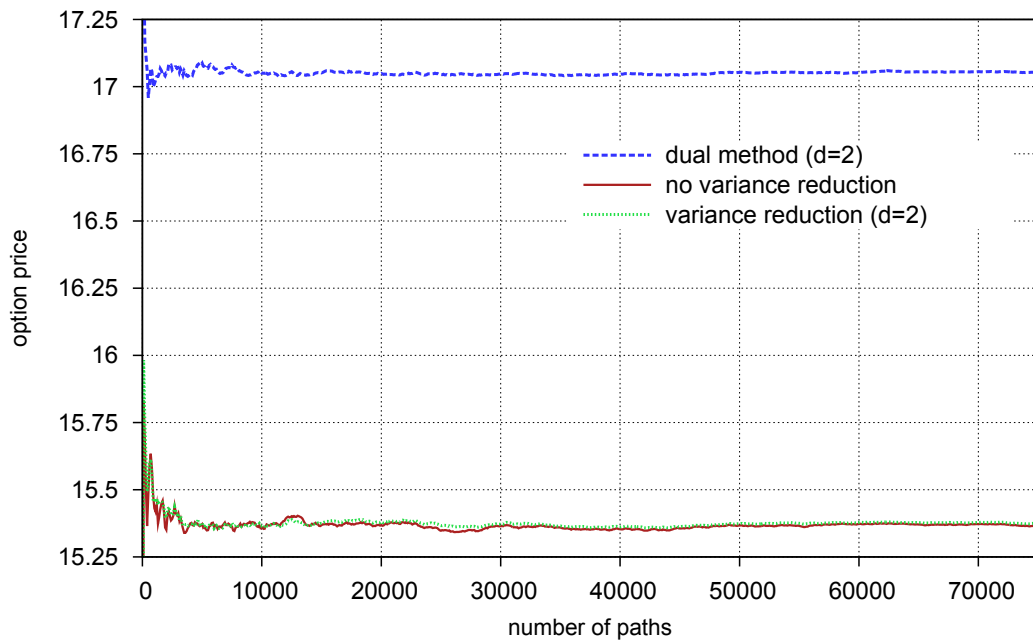
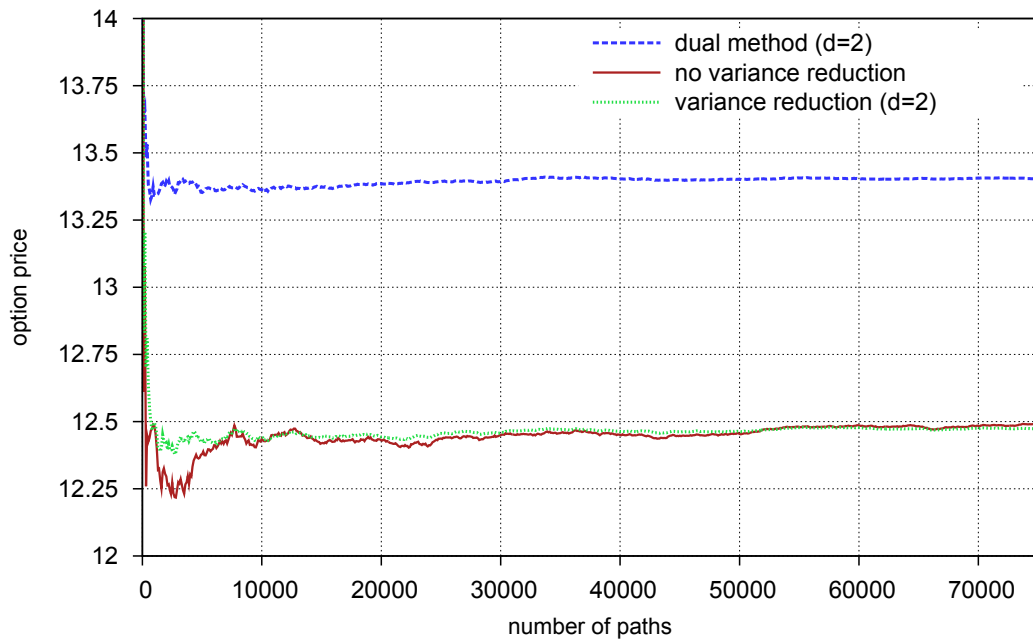


Fig. 4: MC, variance reduced MC, and dual MC for *minimum put option* ( $n = 30, d = 2$ ).  
 Top: high correlation,  $N_{ex} = 10$ ; bottom: low correlation,  $N_{ex} = 100$ .

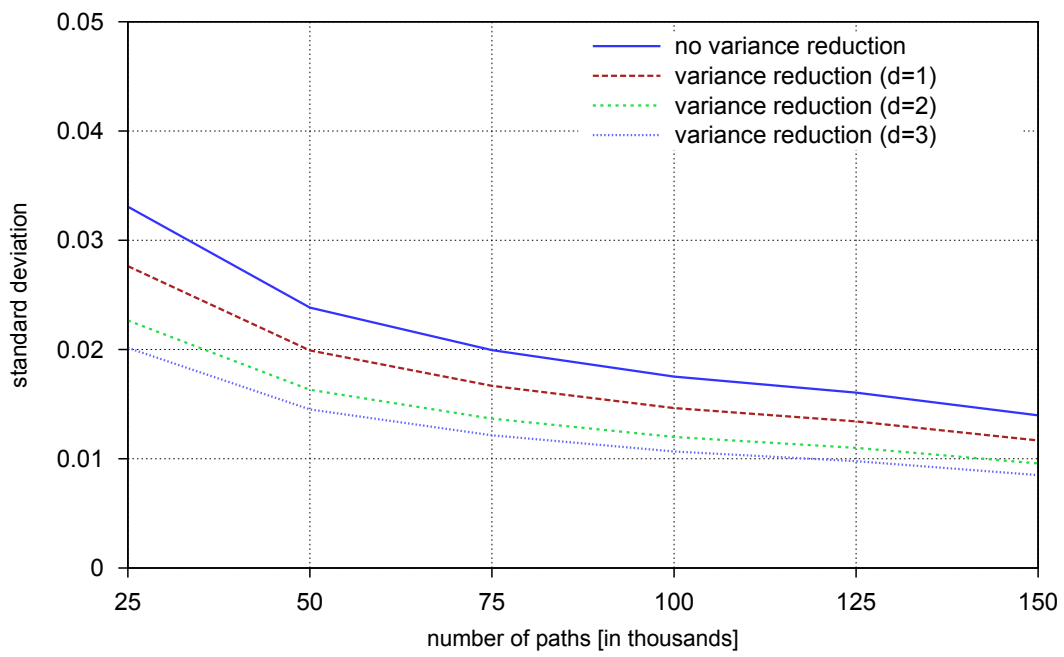
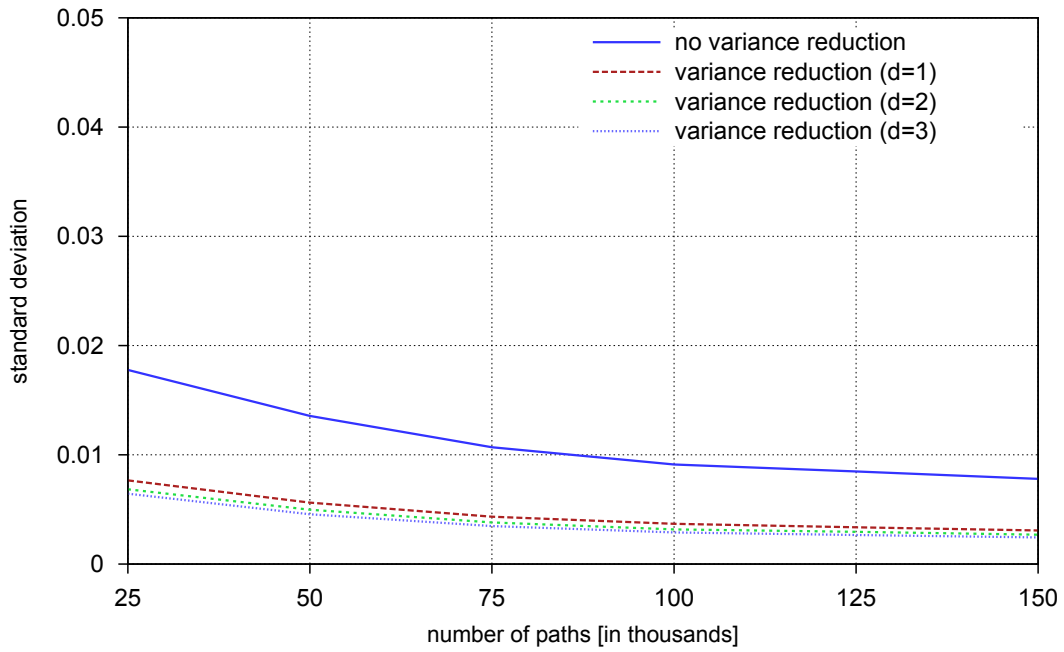


Fig. 5: Standard deviation with and without variance reduction for  $d = 1, 2, 3$  (low correlation,  $n = 30$ ,  $n_{ex} = 10$ ). Top: average put option; bottom: minimum put option.



## REFERENCES

- [1] S. Asmussen and P. W. Glynn. *Stochastic Simulation: Algorithms and Analysis*. Springer, New York, 2007.
- [2] D. Belomestny. Pricing Bermudan options using regression: optimal rates of convergence for lower estimates. *Finance and Stochastics*, 2011.
- [3] D. Belomestny, A. Kolodko, and J. Schoenmakers. Regression methods for stochastic control problems and their convergence analysis. *SIAM J. Control Optim.*, 48(5):3562–3588, 2010.
- [4] P. Carr and D. B. Madan. Option valuation using the fast Fourier transform. *Journal of Computational Finance*, 2(4), 1999.
- [5] C. Chiarella and A. Ziogas. American call options under jump-diffusion processes – A Fourier transform approach. *Applied Mathematical Finance*, 16(1):37–39, 2009.
- [6] R. Cont and P. Tankov. *Financial Modelling with Jump Processes*. Chapman & Hall, Boca Raton, 2004.
- [7] Y. d’Halluin, P. Forsyth, and G. Labahn. A penalty method for American options with jump diffusion processes. *Numer. Math.*, 97(2):321–352, 2004.
- [8] L. Feng and V. Linetsky. Pricing options in jump-diffusion models: An extrapolation approach. *Operations Research*, 56(2):304–325, 2008.
- [9] M. Frigo and S. G. Johnson. The design and implementation of FFTW<sub>3</sub>. *Proceedings of the IEEE*, 93(2):216–231, 2005. Special issue on “Program Generation, Optimization, and Platform Adaptation”.
- [10] G. Fusai and A. Roncoroni. *Implementing Models in Quantitative Finance: Methods and Cases*. Springer, Berlin, 2008.
- [11] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, New York, 2004.
- [12] M. B. Haugh and L. Kogan. Pricing American options: A duality approach. *Operations Research*, 52(2):258–270, 2004.
- [13] P. Heppenger. Option pricing in Hilbert space valued jump-diffusion models using partial integro-differential equations. *SIAM Journal on Financial Mathematics*, 1:454–489, 2010.
- [14] P. Heppenger. Hedging electricity swaptions using partial integro-differential equations. *Stochastic Processes and their Applications*, 122(2):600 – 622, 2012.
- [15] M. Kohler, A. Krzyżak, and N. Todorovic. Pricing of high-dimensional American options by neural networks. *Mathematical Finance*, 20:383–410, 2010.

- [16] A. Kolodko and J. Schoenmakers. Iterative construction of the optimal Bermudan stopping time. *Finance and Stochastics*, 10:27–49, 2006.
- [17] K. Kunisch and S. Volkwein. Galerkin proper orthogonal decomposition methods for parabolic problems. *Numer. Math.*, 90(1):117–148, 2001.
- [18] A. L. Lewis. A simple option formula for general jump-diffusion and other exponential Lévy processes. In *Other Exponential Lévy Processes, Environ Financial Systems and OptionCity.net*, 2001.
- [19] F. A. Longstaff and E. S. Schwartz. Valuing American options by simulation: A simple least-squares approach. *The Review of Financial Studies*, 14:113–147, 2001.
- [20] L. C. G. Rogers. Monte Carlo valuation of American options. *Mathematical Finance*, 12(3):271–286, 2002.
- [21] G. Xu and H. Zheng. Approximate basket options valuation for a jump-diffusion model. *Insurance Math. Econom.*, 45(2):188–194, 2009.