J. Appl. Prob. **54,** 424–443 (2017)
*doi:*10.1017/jpr.2017.9

ON THE CONVERGENCE OF THE QUASI-REGRESSION METHOD: POLYNOMIAL CHAOS AND REGULARITY

JE GUK KIM,* *SungKyunKwan University*

Abstract

We present an analysis of convergence of a quasi-regression Monte Carlo method proposed by Glasserman and Yu (2004). We show that the method surely converges to the true price of an American option even under multiple underlyings via polynomial chaos expansion and weaker conditions than those used in Glasserman and Yu (2004). Further, we show the number of simulation paths grows exponentially in the number of basis functions to obtain convergence in implementing the method. Finally, we propose a rate of convergence considering regularity of value functions.

Keywords: American option; Monte Carlo; polynomial chaos; regularity; quasi-regression

2010 Mathematics Subject Classification: Primary 91G60
Secondary 65C05

1. Introduction

Seeking a closed-form or, at least, an analytical formula for an American option price is a challenging task in practice since it induces an optimal stopping problem to be solved (see, e.g. [9]). Hence, there is an extensive literature on numerical methods for the pricing problem. Among the numerical methods, a least-squares (L-S) Monte Carlo method proposed by Longstaff and Schwartz [12] has been popular with both researchers and practitioners mainly due to its ease of implementation (see [14]). Carriere [2] and Tsisiklis and Van Roy [15] also tackled the pricing problem through Monte Carlo estimation under a dynamic programming formulation although they are different from L-S algorithms, since L-S algorithms use regression only to determine a stopping rule, not to evaluate the value itself. While the implementation of the method is simple it is challenging to conduct an analysis of the convergence of the method.

The first attempt for this challenging work can be found in Clement *et al.* [3] where the authors addressed some results on convergence of the method to an approximation to the true price with a fixed number of basis functions. Later, the authors in [7], [9], and [14] studied the problem under a situation where the number of paths and number of basis functions increase at the same time. Gerhold [7] extended the work in [9] to several Lévy processes. Stentoft [14] resorted to results on series estimators to achieve polynomial growth of the number of paths in the number of basis functions under the assumption that the underlying asset has a bounded state space. In using techniques from statistical learning theory, Egloff [4] and Egloff *et al.* [5] adopted the boundedness assumption for their study on the convergence of the Longstaff and Schwartz method [12]. Zanger [17] improved on the work in [4] and [5] without the boundedness assumption but also through statistical learning theory. However, it should be

Received 17 March 2016; revision received 20 December 2016.
* Postal address: SKK Business School, SungKyunKwan University, 25-2, Sungkyunkwan-ro, Jongno-gu, Seoul 03063, Republic of Korea. Email address: jkim74@vols.utk.edu

noted that Glasserman and Yu [9] and Gerhold [7] analyzed a quasi-regression method different from the standard L-S Monte Carlo regression method proposed by Longstaff and Schwartz in two aspects: the scheme to generate paths and the use of the exact matrix in the calculation of coefficients of basis functions (see [9, p. 2095]).

In the present paper we complement the results on the quasi-regression Monte Carlo method addressed in [9] in a theoretical way. Glasserman and Yu [9] proposed a quasi-regression method and showed that the proposed algorithm converges to an approximation to the true price of the American option with a single underlying asset following Brownian motion, and the number of basis functions (Hermite polynomials) K in a sample size N to obtain convergence is $O(\log N)(O(\sqrt{\log N})$ for geometric Brownian motion using multiples of the powers x^k as basis functions). We show that, with shorter arguments and weaker conditions than those used in [9], the algorithm converges to the true value even under the assumption of multiple underlyings with the critical value $O(\log N)$ through the theory of polynomial chaos expansion for both cases: Brownian (BM) and geometric Brownian motion (GBM). In addition, considering the regularity of the continuation value function, we present a rate of convergence of the algorithm, which may be useful in practice although the rate is not a sharp one.

We must, however, point out that the critical value for GBM itself is not an improvement in comparison to [9]. The result here comes from treating GBM as just a function of BM, which is not costless because moving the exponential function in the option payoff to the payoff will often result in a much larger K. Moreover, the switching between BM and GBM through a transformation here corresponds to a change in basis functions for GBM in [9]; that is, it is a different setting. Therefore, the critical value for GBM here is not comparable to the one for GBM in [9].

In Section 2 we recall the general backward induction framework for pricing of an American option. In Section 3 we assume our market model and introduce some notation for polynomial chaos. The facts needed for deriving the main convergence results are collected there. We propose the algorithm to be analyzed in Section 4. In Section 5 we present two main results: the convergence of the algorithm and a rate of convergence of the algorithm. We will end with some concluding remarks in Section 6.

2. General description of an American option pricing

In this section we present a general framework for pricing of an American option to which our algorithm in the subsequent sections applies. We follow the presentation of [9]. For a more detailed description about the formulation of the framework in many sources, we refer the reader to, e.g. [9].

We assume a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbb{P}$ is the risk-neutral measure. We deal with the problem in a discretized time setting: the option expires in m periods with T as the expiration date and set the early exercise points as $t_0 = 0 < t_1 < \cdots < t_m = T$. Hence, our problem can be considered as an approximation to the price of an American option in discretized time or the exact price of a Bermudan option. A theoretical value $V_{t_n}(x)$ of American option at t_n in state x is given by

$$V_{t_n}(x) = \sup_{\tau \in \Gamma_n} \mathbb{E}[h_\tau(S_\tau) \mid S(t_n), = x],$$

where $h_t \colon \mathbb{R}^d \to \mathbb{R}$ is a payoff function for the option and $h_t(S) \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, Γ_n is the set of all stopping times taking values in $\{t_n, \dots, t_m\}$ adapted to the filtration corresponding to a market model, $S(t) = (S_1(t), \dots, S_d(t))$, where $S(t)$ is a given stochastic process.

https://doi.org/10.1017/jpr.2017.9 Published online by Cambridge University Press

The option value satisfies the backward induction equations

$$V_{t_m}(x) = h_{t_m}(x), \qquad V_{t_n}(x) = \max\{h_{t_n}(x), \mathbb{E}[V_{t_{n+1}}(S(t_{n+1})) \mid S(t_n) = x]\},$$

$n = 0, 1, \dots, m-1$. We can rewrite these with respect to continuation values

$$C^*_{t_n}(x) = \mathbb{E}[V_{t_{n+1}}(S(t_{n+1})) \mid S(t_n) = x], \qquad n = 0, 1, \dots, m-1,$$

as

$$C^*_{t_m}(x) = 0, \qquad C^*_{t_n}(x) = \mathbb{E}[\max\{h_{t_{n+1}}(S(t_{n+1})), C^*_{t_{n+1}}(S(t_{n+1}))\} \mid S(t_n) = x],$$

$n = 0, 1, \dots, m-1$. The option value satisfies

$$V_{t_n}(x) = \max\{h_{t_n}(x), C^*_{t_n}(x)\}.$$

Therefore, we can calculate the value from the continuation values at least from a theoretical perspective. We note that if $S(0)$ is a constant then $C^*_{t_0} = \mathbb{E}[V_{t_1}]$. We further note that deterministic or stochastic discounting can be absorbed into h_{t_n} (see [9]).

3. Preliminaries

First, we assume our market model; the underlying assets $\{S_i\}_{i=1}^d$ follow a correlated GBM with a fixed initial value $S_0 = \{s_1, \dots, s_d\}$ under the risk-neutral measure

$$\mathrm{d}S_i(t) = rS_i(t)\,\mathrm{d}t + \sigma_i S_i(t)\,\mathrm{d}W_i(t), \qquad i = 1, \dots, d,$$

equivalently,

$$S_i(t) = s_i \exp\bigl(\bigl(r - \tfrac{1}{2}\sigma_i^2\bigr)t + \sigma_i W_i(t)\bigr), \qquad i = 1, \dots, d,$$

where $\{W_i\}_{i=1}^d$ is a correlated BM, r is the risk free rate, and σ_i is the volatility of the ith asset. The correlations between W_i and W_j are given by a $d \times d$ matrix ρ with $[\rho]_{ij} = \rho_{ij}$, where $\rho_{ii} = 1$ and $-1 \le \rho_{ij} \le 1$. Since, by the Cholesky decomposition,

$$\rho = HH^\top,$$

where H is a lower triangular matrix,

$$\boldsymbol{W}(t) = H\boldsymbol{Z}(t),$$

where $\boldsymbol{Z}(t)$ is a d-dimensional BM (see [8]). Then, in our discretized setting, for $i = 1, \dots, d$,

$$S_i(t_n) = s_i \exp\biggl(\biggl(r - \frac{1}{2}\sigma_i^2\biggr)t_n + \sigma_i \sum_{j=1}^d h_{ij} Z_j(t_n)\biggr), \qquad n = 0, 1, \dots, m.$$

For notational convenience, we denote $(S_i(t_n))_{i=1}^d$ and $(Z_i(t_n)/\sqrt{t_n})_{i=1}^d$ by $\boldsymbol{S}_n$ and $\boldsymbol{\xi}_n$ for $n = 1, \dots, m$. Note that $\boldsymbol{\xi}_n$ is a random vector consisting of independent and identically distributed (i.i.d.) random variables with a standard normal distribution. The following assumption turns out to be useful.

Assumption 1. *Assume that ρ is positive-definite. Then, since H is invertible, the σ-algebras generated by $\boldsymbol{S}_n$ and $\boldsymbol{\xi}_n$ are equivalent; that is,*

$$\sigma(\boldsymbol{S}_n) = \sigma(\boldsymbol{\xi}_n).$$

https://doi.org/10.1017/jpr.2017.9 Published online by Cambridge University Press

Therefore, by the Doob–Dynkin lemma (see [6]), there exists a Borel-measurable function C_{t_n} from $\mathbb{R}^d$ to $\mathbb{R}$ such that

$$C^*_{t_n}(\boldsymbol{S}_n) = C_{t_n}(\boldsymbol{\xi}_n), \qquad n = 1, \dots, m-1.$$

Thus, for each $n \in \{1, \dots, m-1\}$, we have

$$C^*_{t_n}(\boldsymbol{S}_n) = C_{t_n}(\boldsymbol{\xi}_n) \in L^2(\Omega, \sigma(\boldsymbol{\xi}_n), \mathbb{P}).$$

For notational simplicity, we write C_n for C_{t_n} for each n.

A popular approach for computations involved with random data is to represent them as a polynomial chaos expansion. This method has been applied successfully to research on numerical partial differential equations with uncertain data while the theoretical foundation dates back to the classical Cameron–Martin theorem. We introduce some facts about this method in the context of the current application. We refer the reader to [6] for a detailed development of the theory.

Define the normalized Hermite polynomials $\{\psi_k\}_{k\in\mathbb{N}_0}$ by

$$\psi_k(x) = \frac{1}{\sqrt{k!}} H_k(x), \qquad x \in \mathbb{R},$$

where $H_k(x) = (-1)^k \exp(\frac{1}{2}x^2)(\mathrm{d}^k \exp(-\frac{1}{2}x^2)/\mathrm{d}x^k)$. The normalized Hermite polynomials satisfy

$$\psi'_k(x) = \sqrt{k}\,\psi_{k-1}(x), \qquad k \geq 1, \tag{1}$$

and

$$\psi_{k+1}(x) = \frac{x}{\sqrt{k+1}}\psi_k(x) - \frac{\sqrt{k}}{\sqrt{k+1}}\psi_{k-1}(x), \qquad k \geq 1. \tag{2}$$

We further note that

$$\int_{\mathbb{R}} \psi_m(x)\psi_n(x)\omega(x)\,\mathrm{d}x = \delta_{mn},$$

where ω is the standard normal density function (see [1]).

We consider $L^2(\Omega, \sigma(\xi), \mathbb{P})$, where ξ has a standard normal distribution F_ξ. For any $\varphi \in L^2(\Omega, \sigma(\xi), \mathbb{P})$, by the Doob–Dynkin lemma, there exists a Borel-measurable function $f\colon \mathbb{R} \to \mathbb{R}$ such that $\varphi = f(\xi)$. Then, since the normalized Hermite polynomials constitute an orthonormal system for $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), F_\xi(\mathrm{d}x))$, the set $\{\psi_k(\xi)\}_{k\in\mathbb{N}_0}$ is also an orthonormal system of $L^2(\Omega, \sigma(\xi), \mathbb{P})$. Moreover, Theorems 3.3 and 3.4 in [6] imply that $\{\psi_k(\xi)\}_{k\in\mathbb{N}_0}$ is a complete orthonormal system for $L^2(\Omega, \sigma(\xi), \mathbb{P})$; that is,

$$\varphi = f(\xi) = \sum_{k=0}^{\infty} a_k \psi_k(\xi) \quad \text{in } L^2,$$

where

$$\begin{aligned} a_k &= \langle \varphi, \psi_k(\xi) \rangle \\ &= \int_\Omega \varphi \psi_k(\xi)\,\mathrm{d}\mathbb{P} \\ &= \int_\Omega f(\xi)\psi_k(\xi)\,\mathrm{d}\mathbb{P} \\ &= \int_{\mathbb{R}} f(x)\psi_k(x) F_\xi(\mathrm{d}x), \qquad k \in \mathbb{N}_0. \end{aligned}$$

The above constructing procedure naturally extends to multidimensional cases. Let us consider a random vector $\boldsymbol{\xi}\colon \Omega \to \mathbb{R}^d$, where $\boldsymbol{\xi} = (\xi_1, \dots, \xi_d)$ and $\{\xi_i\}_{i=1}^d$ is i.i.d. with a standard normal distribution. Then, denoting $\{\psi_j^{(i)}\}_{j\in\mathbb{N}_0}$ by the normalized Hermite polynomials corresponding to ξ_i, Theorems 3.6 and 3.7 in [6] imply the set of multivariate tensor products of the polynomials given by

$$\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \prod_{m=1}^{d} \psi_{\alpha_m}^{(m)}(\xi_m), \qquad \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d,$$

is a complete orthonormal system of $L^2(\Omega, \sigma(\boldsymbol{\xi}), \mathbb{P})$. For each $\varphi \in L^2(\Omega, \sigma(\boldsymbol{\xi}), \mathbb{P})$ and a Borel-measurable function $f\colon \mathbb{R}^d \to \mathbb{R}$,

$$\varphi = f(\boldsymbol{\xi}) = \sum_{\boldsymbol{\alpha}\in\mathbb{N}_0^d} a_{\boldsymbol{\alpha}} \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) \quad \text{in } L^2,$$

where

$$a_{\boldsymbol{\alpha}} =< \varphi, \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) >= \int_\Omega \varphi \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\xi})\, \mathrm{d}\mathbb{P} = \int_\Omega f(\boldsymbol{\xi})\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\xi})\, \mathrm{d}\mathbb{P} = \int_{\mathbb{R}^d} f(\boldsymbol{x})\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{x}) F_{\boldsymbol{\xi}}(\mathrm{d}\boldsymbol{x})$$

and $F_{\boldsymbol{\xi}}(\mathrm{d}\boldsymbol{x}) = F_{\xi^{(1)}}(\mathrm{d}x) \times \cdots \times F_{\xi^{(d)}}(\mathrm{d}x)$.

Although the multi-index representation is legitimate in theoretical development, it is impractical to use the multi-index representation for the purpose of a finite-term approximation to a function in $L^2(\Omega, \sigma(\boldsymbol{\xi}), \mathbb{P})$. We therefore introduce a single-index that is more tractable in constructing a finite truncation of an infinite-sum representation of a function in L^2. Among single-index schemes, we adopt the *graded lexicographic order*, which says that higher-degree monomials are bigger and we use lexicographic order to break ties (see Appendix A for an example). By adopting the scheme,

$$\varphi = f(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} a_k \boldsymbol{\psi}_k(\boldsymbol{\xi}),$$

where $a_k =< \varphi, \boldsymbol{\psi}_k(\boldsymbol{\xi}) >,\ k \in \mathbb{N}_0$. Thus, we have

$$C_n(\boldsymbol{\xi}_n) = \sum_{k=0}^{\infty} a_k \boldsymbol{\psi}_k(\boldsymbol{\xi}_n),$$

where $a_k =< C_n(\boldsymbol{\xi}_n), \boldsymbol{\psi}_k(\boldsymbol{\xi}_n) >$. Note that supposing $\boldsymbol{\alpha}(k) = (\alpha_k^1, \dots, \alpha_k^d)$ is the multi-index with $|\boldsymbol{\alpha}(k)| = \sum_{m=1}^d \boldsymbol{\alpha}(k)$ corresponding to k, we have $|\boldsymbol{\alpha}(k)| \le |\boldsymbol{\alpha}(k+1)|$ and $|\boldsymbol{\alpha}(k)| \le k$.

Now, we present an estimate for the fourth moment of Hermite polynomials useful in developing our main results later.

Proposition 1. *Let ψ and $\boldsymbol{\psi}$ be Hermite polynomials and multidimensional Hermite polynomials, respectively. Then, the following hold:*

(i) *for sufficiently large k, there exist positive constants C and $\tilde{C}$ such that*

$$C\frac{3^{2k}}{k} \le \mathbb{E}[\psi_k^4] \le \tilde{C}\frac{3^{2k}}{k};$$

(ii) *for $k \ge 1$, there exists a positive constant C such that*

$$\mathbb{E}[\boldsymbol{\psi}_k^4] \le C3^{2|\boldsymbol{\alpha}(k)|}.$$

https://doi.org/10.1017/jpr.2017.9 Published online by Cambridge University Press

Proof. From Theorem 2.1. in [12], it is obvious that (i) holds. For (ii), we note that

$$\mathbb{E}[\boldsymbol{\psi}_k^4] = \prod_{j=1}^{d} \mathbb{E}[\psi^4_{\boldsymbol{\alpha}(k)_j}].$$

Then, by (i), we have $\mathbb{E}[\boldsymbol{\psi}_k^4] \leq C3^{2|\boldsymbol{\alpha}(k)|}$, which completes the proof. □

Note that C denotes a generic positive constant throughout the paper.

4. Algorithm

We recall the quasi-regression algorithm to be analyzed that was proposed in [10].
Step 1. Set $\hat{C}_m = 0$ and $\hat{V}_m = \max\{h_m, \hat{C}_m\} = h_m$.
Step 2. For each $n = 1, \dots, m-1$, starting from $m-1$, we repeat the following:

- generate N independent copies $\{\boldsymbol{S}_1^i, \dots, \boldsymbol{S}_{n+1}^i\}$ of path $\{\boldsymbol{S}_1, \dots, \boldsymbol{S}_{n+1}\}$, $i = 1, \dots, N$, up to time t_{n+1}, independent of all previously generated paths. Set

$$\hat{\gamma}_{n,k} = \frac{1}{N}\sum_{i=1}^{N} \hat{V}_{n+1}(\boldsymbol{S}_{n+1}^i)\boldsymbol{\psi}_{n,k}(\boldsymbol{S}_n^i), \qquad k = 0, \dots, K,$$

  calculate the coefficients $\hat{\beta}_n = \Psi_n^{-1}\hat{\gamma}_n$, where K represents the number of basis functions and Ψ_n is defined below, and set

$$\hat{C}_n = \sum_{k=0}^{K} \hat{\beta}_{n,k}\boldsymbol{\psi}_{n,k} \quad \text{and} \quad \hat{V}_n = \max\{h_n, \hat{C}_n\}.$$

Step 3. Set $\hat{C}_{N,K,0}(\boldsymbol{S}_0) = (1/N)\sum_{i=1}^{N} \hat{V}_1(\boldsymbol{S}_1^i)$ and $\hat{V}_0(\boldsymbol{S}_0) = \max\{h_0(\boldsymbol{S}_0), \hat{C}_{N,K,0}(\boldsymbol{S}_0)\}$.
In this algorithm, $\boldsymbol{S}_0$ is fixed and the ψ' are general basis functions . We note that step 2 is different from the algorithm in [14], which generates a single set of paths for all dates. Moreover, we also note that the present algorithm has another feature different from the algorithm in [13]; in the regression process, we use the exact matrix

$$\Psi_n = \mathbb{E}[\psi_n(S_n)\psi_n(S_n)^\top]$$

instead of its sample counterpart

$$\frac{1}{N}\sum_{i=1}^{N} \psi_n(S_n^i)\psi_n(S_n^i)^\top,$$

calculated from the simulated values themselves; the method adopting this alteration is called the quasi-regression method (see [10, pp. 2094–2095] or [13] for a detailed description of the method).

Our purpose is to analyze convergence of the algorithm for two concrete examples: BM and GBM. To this end we alter step 2 in the algorithm to be more convenient for our purpose. Specifically, we choose a proper transformation ϕ so as to have $\phi(\boldsymbol{S}) = \boldsymbol{\xi}$ and take the composite $\psi \circ \phi$ as basis functions in the algorithm, where the ψ' are Hermite polynomials.

Considering the assumption and results in Section 3 it is possible for one to have this kind of basis functions at least for correlated BM and GBM under Assumption 1. For instance, in the one-dimensional case, the basis functions for GBM are, by the facts from Section 3 and the argument in the present section,

$$\psi_k(\phi_{t_n}(S_{t_n})) = \psi_k(\xi_{t_n}),$$

where $\phi_n(x) = (\log x + t_n/2)/\sqrt{t_n}$. The resulting modified version of step 2 is as follows.

Step 2. For each $n = 1, \dots, m-1$, starting from $m-1$, we repeat the following:

- generate N independent copies $\{\boldsymbol{Z}_1^i, \dots, \boldsymbol{Z}_{n+1}^i\}$ of path $\{\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_{n+1}\}$, $i = 1, \dots, N$, up to time t_{n+1}, independent of all previously generated paths. Calculate

$$\boldsymbol{S}_{n+1}^i \text{ and } \boldsymbol{\xi}_n^i$$

and

$$\hat{\beta}_{n,k} = \frac{1}{N}\sum_{i=1}^{N} \hat{V}_{n+1}(\boldsymbol{S}_{n+1}^i)\boldsymbol{\psi}_k(\boldsymbol{\xi}_n^i), \qquad k = 0, \dots, K. \tag{3}$$

- Set

$$\hat{C}_{N,K,n} = \sum_{k=0}^{K} \hat{\beta}_{n,k}\boldsymbol{\psi}_k \quad \text{and} \quad \hat{V}_n = \max\{h_n, \hat{C}_{N,K,n}\}. \tag{4}$$

The meaning of the additional subindices N and K will be clear in the next section.

We note that Glasserman and Yu [10] used the expectation of the weighted L^2-norm on functions $G\colon \mathbb{R} \to \mathbb{R}$ that slightly differs from the ordinary L^2-norm (see [10, p. 2106]). We alter slightly step 3 so as to use the ordinary L^2-norm to estimate errors in the analysis of the convergence of the algorithm. For $n = 0$, generate N independent copies $\boldsymbol{S}_1^i$ of $\boldsymbol{S}_1$ independent of all previously generated paths. With these samples, we calculate $\hat{C}_{N,K,0}(\boldsymbol{S}_0)$. The overhead for this additional computational effort is negligible. Moreover, at the cost of adding this step, we gain a huge reward; there are many assumptions in [10] but those assumptions are unnecessary, which will be clear in the next section.

Before going to the main results of this paper, we now reconsider an assumption for the single-period problem in [10] where the dimension of the underlying asset is 1. Glasserman and Yu in [10] proposed three assumptions (A1), (A2), and (A3) to obtain the desired result for the single-period problem:

(A1) $|\beta| = 1$;

(A2) $h_2(S_{t_2}) = \sum_{k=0}^{K} a_k \psi_{2k}(S_{t_2})$ for some constants a_k;

(A3) $\psi_{nk}(S_n)$ are martingales, up to a deterministic function of time.

Among them, a remark on assumption (A2) is needed. Specifically, we note that

$$h_2(S_{t_2}) = H(\xi_{t_2})$$

for some Borel-measurable function $H\colon \mathbb{R} \to \mathbb{R}$. Hence, by the polynomial chaos expansion in Section 4, we have

$$h_2(S_{t_2}) = \sum_{k=0}^{\infty} a_k \psi_k(\xi_{t_2}),$$

which is the motivating idea for the present section.

Gerhold [8] gave some intuitive justification of the infinite series representation above and the interpretation of (A2) as a good approximation of the payoff at t_2 (see Gerhold [7, p. 596]). However, in our setting, the intuitive justification turns into a rigorous one.

5. Main results

In this section we present two main results: the convergence of the algorithm when multiple underlying assets are considered and a rate of convergence of the algorithm. To this end we introduce several artificial devices useful for the proof of the main results below. Define, for $n \in \{1, 2, \dots, m-1\}$,

$$C_{K,n} = P_K V_{n+1} = \sum_{k=0}^{K} \beta_{n,k} \boldsymbol{\psi}_k, \tag{5}$$

where P_K is the orthogonal projection onto $\operatorname{span}\{\psi_0, \dots, \psi_K\}$ and

$$\beta_{n,k} = \mathbb{E}[V_{n+1}\psi_k]. \tag{6}$$

Define an approximation to the backward induction equations as follows: $\bar{V}_m = h_m$,

$$\bar{C}_{K,n} = P_K \bar{V}_{n+1} = \sum_{k=0}^{K} \bar{\beta}_{n,k} \psi_k, \qquad n \in \{1, \dots, m-1\}, \tag{7}$$

where $\bar{V}_{n+1} = \max\{h_{n+1}, \bar{C}_{n+1}\}$ and

$$\bar{\beta}_{n,k} = \mathbb{E}[\bar{V}_{n+1}\boldsymbol{\psi}_k]. \tag{8}$$

Finally, define, for $n \in \{1, \dots, m-1\}$,

$$\tilde{C}_{N,K,n} = \sum_{k=0}^{K} \tilde{\beta}_{n,k} \boldsymbol{\psi}_k, \tag{9}$$

where

$$\tilde{\beta}_{n,k} = \frac{1}{N} \sum_{i=1}^{N} \bar{V}_{n+1}(\boldsymbol{S}_{n+1}^{i}) \boldsymbol{\psi}_k(\boldsymbol{\xi}_n^{i}). \tag{10}$$

Now, we address the single period problem where $m = 2$. We introduce an assumption needed to derive the main results.

Assumption 2. *Assume that $\mathbb{E}[h_n^4] < \infty$ for each n.*

The Assumption 2 is remarkably less restrictive than the ones for the fourth moment of h in the literature (see [10, Equation (B3)] and [8, Theorem 6]). We now address the result for the case of the single-period and single underlying asset.

Theorem 1. (i) *If m=2 and $K = ((1-\delta)/c)\log N$, where $c = \log 3^2$ and $\delta \in (0, 1)$, $\hat{C}_{N,K,0}(S_0)$ converges to $C_0(S_0)$ in L^2 as $N \to \infty$.*

(ii) *If $K = ((1+\delta)/c)\log N$, $\delta > 0$, the algorithm, for some payoff function, diverges to the infinite in L^2 as $N \to \infty$.*

https://doi.org/10.1017/jpr.2017.9 Published online by Cambridge University Press

Proof. Note that, for the definitions of $\hat{\beta}_{1,k}$, $\hat{C}_{N,K,1}$, $C_{K,1}$, and $\beta_{1,k}$, see (3)–(6), respectively.

(i) First, we estimate $\mathbb{E}[(\hat{C}_{N,K,1}(\xi_1) - C_{K,1}(\xi_1))^2]$ under the assumption of the algorithm that the coefficients of $\hat{C}_{N,K,1}$ and $C_{K,1}$ are independent of ξ_1.

By independence and orthogonality,

$$\mathbb{E}[(\hat{C}_{N,K,1}(\xi_1) - C_{K,1}(\xi_1))^2] = \mathbb{E}[\sum_{k=0}^{K}(\hat{\beta}_{1,k} - \beta_{1,k})^2].$$

Since $\mathbb{E}[\hat{\beta}_{1,k}] = \beta_{1,k}$ since $\hat{V}_2 = h_2 = V_2$,

$$\sum_{k=0}^{K}\mathbb{E}[(\hat{\beta}_{1,k} - \beta_{1,k})^2] = \frac{1}{N}\sum_{k=0}^{K}\operatorname{var}(\psi_{1,k}(\xi_1)h_{t_2}(S_2)) \le \frac{1}{N}\sum_{k=0}^{K}\mathbb{E}[\psi_{1,k}^2(\xi_1)h_{t_2}^2(S_2)].$$

Thus, by the Cauchy–Schwarz inequality and Proposition 1,

$$\mathbb{E}\bigg[\sum_{k=0}^{K}(\hat{\beta}_{1,k} - \beta_{1,k})^2\bigg] \le C\frac{(K+1)3^{2K}}{N}.$$

Therefore, since $C_{K,1} \to C_1$ in L^2, by the triangle inequality, we have $\hat{C}_{N,K,1} \to C_1$ in L^2. Now, we show that $\hat{C}_{N,K,0}(S_0)$ converges to $C_0(S_0) = \mathbb{E}[V_1]$ in L^2. Note that

$$\begin{aligned}
&\mathbb{E}[(\hat{C}_0(S_0) - \mathbb{E}[V_1])^2] \\
&\quad = \mathbb{E}\bigg[\bigg(\frac{1}{N}\sum_{i=1}^{N}\hat{V}_1(S_1^i) - \mathbb{E}[V_1]\bigg)^2\bigg] \\
&\quad \le 2\mathbb{E}\bigg[\bigg(\frac{1}{N}\sum_{i=1}^{N}\hat{V}_1(S_1^i) - \frac{1}{N}\sum_{i=1}^{N}V_1(S_1^i)\bigg)^2\bigg] + 2\mathbb{E}\bigg[\bigg(\frac{1}{N}\sum_{i=1}^{N}V_1(S_1^i) - \mathbb{E}[V_1]\bigg)^2\bigg] \\
&\quad \le 2\mathbb{E}[(\hat{C}_{N,K,1}(\xi_1^i) - C_1(\xi_1^i))^2] + 2\frac{\operatorname{var}(V_1)}{N}.
\end{aligned}$$

Then, since the coefficients of $\hat{C}_{N,K,1}$ are independent of ξ_1^i by alteration of step 3, we obtain the convergence as $N \to \infty$.

(ii) It is enough to address an example showing the divergence.

Let $h_{t_2}(S_2) = (t_2/t_1)^{K/2}\psi_{2K}(\xi_2)$ via ϕ. By the triangle inequality and the fact that $C_{K,1} \to C_1$ in L^2, it is sufficient to show that $\mathbb{E}[(\hat{C}_{N,K,1}(\xi_1) - C_{K,1}(\xi_1))^2]$ diverges to ∞. Note that

$$\begin{aligned}
&\mathbb{E}[(\hat{C}_{N,K,1}(\xi_1) - C_{K,1}(\xi_1))^2] \\
&\quad = \mathbb{E}\bigg[\sum_{k=0}^{K}(\hat{\beta}_{1,k} - \beta_{1,k})^2\bigg] \\
&\quad = \sum_{k=0}^{K}\operatorname{var}(\hat{\beta}_k) \\
&\quad = \frac{1}{N}\sum_{k=0}^{K}\mathbb{E}\bigg[\bigg(\frac{t_2}{t_1}\bigg)^K\psi_{2K}^2(\xi_2)\psi_{1k}^2(\xi_1)\bigg] - \frac{1}{N}\sum_{k=0}^{K}\bigg(\mathbb{E}\bigg[\bigg(\frac{t_2}{t_1}\bigg)^{K/2}\psi_{2K}(\xi_2)\psi_{1k}(\xi_1)\bigg]\bigg)^2.
\end{aligned}$$

https://doi.org/10.1017/jpr.2017.9 Published online by Cambridge University Press

Then, by [10, Equation (28)],

$$\begin{aligned}\mathbb{E}[(\hat{C}_{N,K,1}(\xi_1) - C_{K,1}(\xi_1))^2] &= \mathbb{E}[\sum_{k=0}^{K}(\hat{\beta}_{1,k} - \beta_{1,k})^2] \\ &= \frac{1}{N}\sum_{k=0}^{K}\mathbb{E}\bigg[\bigg(\frac{t_2}{t_1}\bigg)^K \psi_{2K}^2(\xi_2)\psi_{1k}^2(\xi_1)\bigg] - \frac{1}{N} \\ &\geq \frac{1}{N}\mathbb{E}\bigg[\bigg(\frac{t_2}{t_1}\bigg)^K \psi_{2K}^2(\xi_2)\psi_{1K}^2(\xi_1)\bigg] - \frac{1}{N}.\end{aligned}$$

Now, we note that $\mathbb{E}[t_2^{K/2}\sqrt{K!}\,\psi_{2K}(\xi_2) \mid \xi_1] = t_1^{K/2}\sqrt{K!}\,\psi_{1K}(\xi_1)$ (see [10, pp. 2098–2099]). Then, by the Jensen inequality,

$$\begin{aligned}\mathbb{E}[(\hat{C}_{N,K,1}(\xi_1) - C_{K,1}(\xi_1))^2] &= \mathbb{E}\bigg[\sum_{k=0}^{K}(\hat{\beta}_{1,k} - \beta_{1,k})^2\bigg] \\ &\geq \frac{1}{N}\mathbb{E}\bigg[\bigg(\frac{t_2}{t_1}\bigg)\psi_{2K}^2(\xi_2)\psi_{1K}^2(\xi_1)\bigg] - \frac{1}{N} \\ &= \frac{1}{N}\mathbb{E}\bigg[\frac{1}{K!}\bigg(\frac{1}{t_1}\bigg)^K \psi_{1K}^2(\xi_1)\mathbb{E}[(t_2^{K/2}\sqrt{K!}\,\psi_{2K}(\xi_2))^2 \mid \xi_1]\bigg] - \frac{1}{N} \\ &\geq \frac{1}{N}\mathbb{E}[\psi_{1K}^4(\xi_1)] - \frac{1}{N}.\end{aligned}$$

Finally, by Proposition 1, we have

$$\mathbb{E}[(\hat{C}_{N,K,1}(\xi_1) - C_{K,1}(\xi_1))^2] \geq C\frac{3^{2K}}{NK} - \frac{1}{N},$$

which completes the proof. □

Remark 1. Given $L \in \mathbb{N}_0$, consider a set $S = \{\boldsymbol{\psi}_k \colon |\alpha(k)| \leq L\}$. We note that $|S| = (d+L)!/d!\,L!$. Thus, $K \leq (d + |\alpha(K)|)!/d!\,|\alpha(K)|!$. With

$$h_{t_2}(\boldsymbol{S}_2) = \bigg(\frac{t_2}{t_1}\bigg)^{\sum_j^d(\boldsymbol{\alpha}(K)_j)/2^d}\boldsymbol{\psi}_{2K}(\boldsymbol{\xi}_2),$$

the proof of the theorem also holds for the case of multiple underlying assets by the independence of multidimensional Hermite polynomials with a generalized result $|\boldsymbol{\alpha}(K)| = O(\log N)$. When $d = 1$, the result is exactly same as the one in the above theorem.

Remark 2. Glasserman and Yu, after [10, Theorem 1] dealing with the convergence for single-period problem, stated: 'This result shows rather precisely that, from a sample size of N, the highest K for which coefficients of polynomials of order K can be estimated uniformly well is $O(\log N)$.' Now, we state that [10, Theorem 1] shows precisely that the sample size N required to achieve convergence uniformly well over payoff functions grows exponentially in K.

We now extend this result to the case of multi-period problems. To do so, we need two lemmas.

Lemma 1. (i) *For* $n = m-1$,

$$\sum_{k=0}^{K} \mathbb{E}[(\hat{\beta}_{m-1,k} - \tilde{\beta}_{m-1,k})^2 = 0.$$

(ii) *For each* $n \in \{1, \dots, m-2\}$,

$$\sum_{k=0}^{K} \mathbb{E}[(\hat{\beta}_{n,k} - \tilde{\beta}_{n,k})^2] \le 2^{m-n-1} A_K^{m-n-1} \sum_{l=1}^{m-n-1} \sum_{k=0}^{K} \mathbb{E}[(\tilde{\beta}_{m-l,k} - \bar{\beta}_{m-l,k})^2],$$

where $A_K = (K+1)^2 \max_{0\le k\le K} \mathbb{E}[\boldsymbol{\psi}_k^4]$.

Proof. See Appendix A. □

Lemma 2. *For each* $n \in \{1, \dots, m-1\}$,

$$\sum_{k=0}^{K} \mathbb{E}[(\tilde{\beta}_{n,k} - \bar{\beta}_{n,k})^2] \le C \frac{(K+1)^3}{N} \Big(\max_{0\le k\le K} \sqrt{\mathbb{E}[\boldsymbol{\psi}_k^4]} + \max_{0\le k\le K} \mathbb{E}[\boldsymbol{\psi}_k^4] \Big).$$

Proof. See Appendix A. □

We are now in position to address the result for the case of multiple underlying assets.

Theorem 2. *Suppose that Assumptions 1 and 2 hold. Then, if* $K = ((1-\delta)/c)\log N$, *where* $c = \log 3^{2m}$ *and* $\delta \in (0,1)$, *the algorithm converges to the true value of an American option in* L^2 *as* $N \to \infty$.

Proof. Note that for each $n \in \{1, \dots, m-1\}$,

$$\begin{aligned}
&\mathbb{E}[(\hat{C}_{N,K,n}(\boldsymbol{\xi}_n) - C_n(\boldsymbol{\xi}_n))^2] \\
&\quad \le 4(\mathbb{E}[(\hat{C}_{N,K,n}(\boldsymbol{\xi}_n) - \tilde{C}_{N,K,n}(\boldsymbol{\xi}_n))^2] + \mathbb{E}[(\tilde{C}_{N,K,n}(\boldsymbol{\xi}_n) - \bar{C}_{K,n}(\boldsymbol{\xi}_n))^2] \\
&\qquad + \mathbb{E}[(\bar{C}_{K,n}(\boldsymbol{\xi}_n) - C_{K,n}(\boldsymbol{\xi}_n))^2] + \mathbb{E}[(C_{K,n}(\boldsymbol{\xi}_n) - C_n(\boldsymbol{\xi}_n))^2]).
\end{aligned}$$

For the definitions of $\hat{\beta}_{n,k}$, $\hat{C}_{N,K,n}$, $C_{K,n}$, $\beta_{n,k}$, $\bar{C}_{K,n}$, $\bar{\beta}_{n,k}$, $\tilde{C}_{N,K,n}$, and $\tilde{\beta}_{1,k}$ see (3)–(10), respectively. We estimate each term on the right-hand-side of the above inequality under the assumption of the algorithm that the coefficients of $\hat{C}_{N,K,n}$ and $\tilde{C}_{N,K,n}$ are independent of $\boldsymbol{\xi}_n$:
(i) For $\mathbb{E}[(\hat{C}_{N,K,n}(\boldsymbol{\xi}_n) - \tilde{C}_{N,K,n}(\boldsymbol{\xi}_n))^2]$. Note that, by independence,

$$\begin{aligned}
\mathbb{E}[(\hat{C}_{N,K,n}(\boldsymbol{\xi}_n) - \tilde{C}_{N,K,n}(\boldsymbol{\xi}_n))^2] &= \mathbb{E}\bigg[\bigg(\sum_{k=0}^{K} (\hat{\beta}_{n,k} - \tilde{\beta}_{n,k}) \psi_k(\boldsymbol{\xi}_n)\bigg)^2\bigg] \\
&= \mathbb{E}\bigg[\sum_{k=0}^{K} (\hat{\beta}_{n,k} - \tilde{\beta}_{n,k})^2\bigg].
\end{aligned}$$

Then, by Lemmas 1 and 2 and Proposition 1,

$$\begin{aligned}
&\mathbb{E}[(\hat{C}_{N,K,n}(\boldsymbol{\xi}_n) - \tilde{C}_{N,K,n}(\boldsymbol{\xi}_n))^2] \\
&\quad \le 2^{m-n-1} A_K^{m-n-1} \sum_{l=1}^{m-n-1} \sum_{k=0}^{K} \mathbb{E}[(\tilde{\beta}_{m-l,k} - \bar{\beta}_{m-l,k})^2]
\end{aligned}$$

$$
\begin{aligned}
&\le C\frac{A_K^{m-n-1}(K+1)^3}{N}\Big(\max_{0\le k\le K}\sqrt{\mathbb{E}[\boldsymbol{\psi}_k^4(\boldsymbol{\xi})]}+\max_{0\le k\le K}\mathbb{E}[\boldsymbol{\psi}_k^4(\boldsymbol{\xi})]\Big)\\
&\le C\frac{(K+1)^{2m-2n+1}}{N}3^{2(m-n)K}\\
&\le C\frac{(K+1)^{2m}3^{2m\,|\boldsymbol{\alpha}(K)|}}{N},
\end{aligned}
$$

where C is a generic positive constant.

(ii) For $\mathbb{E}[(\tilde{C}_{N,K,n}(\boldsymbol{\xi}_n)-\bar{C}_{K,n}(\boldsymbol{\xi}_n))^2]$. Since

$$
\mathbb{E}[(\tilde{C}_{N,K,n}(\boldsymbol{\xi}_n)-\bar{C}_{K,n}(\boldsymbol{\xi}_n))^2]=\mathbb{E}\Bigg[\sum_{k=0}^{K}(\tilde{\beta}_{n,k}-\bar{\beta}_{n,k})^2\Bigg],
$$

by Lemmas 1, 2 and Proposition 1,

$$
\begin{aligned}
&\mathbb{E}[(\tilde{C}_{N,K,n}(\boldsymbol{\xi}_n)-\bar{C}_{K,n}(\boldsymbol{\xi}_n))^2]\\
&\quad\le C(K+1)^3\frac{1}{N}\Big(\max_{0\le k\le K}\sqrt{\mathbb{E}[\boldsymbol{\psi}_k^4(\boldsymbol{\xi})]}+\max_{0\le k\le K}\mathbb{E}[\boldsymbol{\psi}_k^4(\boldsymbol{\xi})]\Big)\\
&\quad\le C\frac{(K+1)^3 3^{2\,|\boldsymbol{\alpha}(K)|}}{N}.
\end{aligned}
$$

(iii) For $\mathbb{E}[(\bar{C}_{K,n}(\boldsymbol{\xi}_n)-C_{K,n}(\boldsymbol{\xi}_n))^2]$ and $\mathbb{E}[(C_{K,n}(\boldsymbol{\xi}_n)-C_n(\boldsymbol{\xi}_n))^2]$. Note that

$$
\begin{aligned}
\mathbb{E}[(\bar{C}_{K,n}(\boldsymbol{\xi}_n)-C_{K,n}(\boldsymbol{\xi}_n))^2]&=\sum_{k=0}^{K}(\bar{\beta}_{n,k}-\beta_{n,k})^2\\
&=\sum_{k=0}^{K}(\mathbb{E}[(\bar{V}_{n+1}(\boldsymbol{S}_{n+1})-V_{n+1}(\boldsymbol{S}_{n+1}))\boldsymbol{\psi}_k(\boldsymbol{\xi}_n)])^2.
\end{aligned}
$$

Then

$$
\begin{aligned}
\mathbb{E}[(\bar{C}_{K,n}(\boldsymbol{\xi}_n)-C_{K,n}(\boldsymbol{\xi}_n))^2]&=\mathbb{E}[(P_K(\bar{V}_{n+1}(\boldsymbol{S}_{n+1})-V_{n+1}(\boldsymbol{S}_{n+1})))^2]\\
&\le \mathbb{E}[(\bar{V}_{n+1}(\boldsymbol{S}_{n+1})-V_{n+1}(\boldsymbol{S}_{n+1}))^2]\\
&\le \mathbb{E}[|\max\{h_{n+1}(\boldsymbol{S}_{n+1}),\bar{C}_{K,n+1}(\boldsymbol{\xi}_{n+1})\}\\
&\qquad -\max\{h_{n+1}(\boldsymbol{S}_{n+1}),C_{n+1}(\boldsymbol{\xi}_{n+1})\}|^2]\\
&\le \mathbb{E}[(\bar{C}_{K,n+1}(\boldsymbol{\xi}_{n+1})-C_{n+1}(\boldsymbol{\xi}_{n+1}))^2]\\
&\le 2\mathbb{E}[(\bar{C}_{K,n+1}(\boldsymbol{\xi}_{n+1})-C_{K,n+1}(\boldsymbol{\xi}_{n+1}))^2]\\
&\quad +2\mathbb{E}[(C_{K,n+1}(\boldsymbol{\xi}_{n+1})-C_{n+1}(\boldsymbol{\xi}_{n+1}))^2].
\end{aligned}
$$

Thus, by repeating the procedure, we have

$$
\begin{aligned}
\mathbb{E}[(\bar{C}_{K,n}(\boldsymbol{\xi}_n)-C_{K,n}(\boldsymbol{\xi}_n))^2]&\le 2^{m-n-1}\mathbb{E}[(\bar{C}_{K,m-1}(\boldsymbol{\xi}_{m-1})-C_{K,m-1}(\boldsymbol{\xi}_{m-1}))^2]\\
&\quad+\sum_{l=1}^{m-n-1}2^{m-n-l}\mathbb{E}[(C_{K,m-l}(\boldsymbol{\xi}_{m-l})-C_{m-l}(\boldsymbol{\xi}_{m-l}))^2].
\end{aligned}
$$

https://doi.org/10.1017/jpr.2017.9 Published online by Cambridge University Press

Since $\bar{V}(\boldsymbol{S}_m) = h_m(\boldsymbol{S}_m) = V_m(\boldsymbol{S}_m)$, we have

$$\mathbb{E}[(\bar{C}_{K,n}(\boldsymbol{\xi}_n) - C_{K,n}(\boldsymbol{\xi}_n))^2] \leq 2^{m-n-1} \sum_{l=1}^{m-n-1} \mathbb{E}[(C_{K,m-l}(\boldsymbol{\xi}_{m-l}) - C_{m-l}(\boldsymbol{\xi}_{m-l}))^2].$$

Thus,

$$\begin{aligned} &\mathbb{E}[(\bar{C}_{K,n}(\boldsymbol{\xi}_n) - C_{K,n}(\boldsymbol{\xi}_n))^2] + \mathbb{E}[(C_{K,n}(\boldsymbol{\xi}_n) - C_n(\boldsymbol{\xi}_n))^2] \\ &\quad \leq 2^{m-n-1} \sum_{l=1}^{m-n} \mathbb{E}[(C_{K,m-l}(\boldsymbol{\xi}_{m-l}) - C_{m-l}(\boldsymbol{\xi}_{m-l}))^2]. \end{aligned}$$

By (i)–(iii), we finally reach

$$\begin{aligned} &\mathbb{E}[(\hat{C}_{N,K,n}(\boldsymbol{\xi}_n) - C_n(\boldsymbol{\xi}_n))^2] \\ &\quad \leq C_1 \frac{(K+1)^{2m} 3^{2m|\boldsymbol{\alpha}(K)|}}{N} + C_2 \sum_{l=1}^{m-n} \mathbb{E}[(C_{K,m-l}(\boldsymbol{\xi}_{m-l}) - C_{m-l}(\boldsymbol{\xi}_{m-l}))^2]. \end{aligned}$$

Therefore, as $N \to \infty$, $\hat{C}_{N,K,n}(\boldsymbol{\xi}_n)$ converges to $C_n(\boldsymbol{\xi}_n)$ in L^2 for each $n \in \{1, \ldots, m-1\}$, which completes the proof. □

Remark 3. We make several remarks about the theorem. For the case of multi-periods and a single underlying asset, we observe that the continuation value function at t_{m-1} is the same as the one in the single-period problem. Thus, the critical relation $O(\log N)$ for the single-period problem still holds for the multi-periods problem. Since the observation is also true for the case of multiple underlying assets by Remark 1, the critical relation $O(\log N)$ still holds for this case. Therefore, for any case, the critical relation is $O(\log N)$ for GBM. Furthermore, we note that the proof still holds for correlated BM by using the proper transformation ϕ. Therefore, we conclude that the highest K to achieve convergence is $O(\log N)$ for any case. However, this result for GBM is not comparable to the one for GBM in [9] as mentioned in the introduction.

Next, we present a rate of convergence of the algorithm considering the regularity of the continuation value function C_n. To this end, we need some notation and definitions about regularity for use in the next section. First, we introduce some basic notation for regularity which is common in partial differential equations. We follow [16, Section 2] and [11]. Let $\Lambda = \{x \mid -\infty < x < \infty\}$ and $\omega(x) = (1/\sqrt{2\pi}) \exp(-\frac{1}{2}x^2)$. Define

$$L^2_\omega(\Lambda) = \{v \mid v \text{ is measurable and } \|v\|_{L^2_\omega(\Lambda)} < \infty\},$$

where $\|v\|_{L^2_\omega(\Lambda)} = (\int_\Lambda |v(x)|^2 \omega(x)\, \mathrm{d}x)^{1/2}$. Further, let $\partial_x v = \partial v/\partial x$, and, for a nonnegative integer r,

$$H^r_\omega(\Lambda) = \{v \mid \partial^k_x v \in L^2_\omega(\Lambda),\ 0 \leq k \leq r\}.$$

The semi-norm and the norm of $H^r_\omega(\Lambda)$ are given by

$$|v|_{H^r_\omega(\Lambda)} = \|\partial^r_x v\|_{L^2_\omega(\Lambda)} \quad \text{and} \quad \|v\|_{H^r_\omega(\Lambda)} = \left(\sum_{k=0}^{r} |v|^2_{H^k_\omega(\Lambda)}\right)^{1/2}.$$

Similarly, for d-dimensions, let

$$\Lambda_i = \{x_i \mid -\infty < x_i < \infty\}, \qquad \Lambda^d = \Lambda_1 \times \Lambda_2 \times \cdots \times \Lambda_d, \qquad x = (x_1, x_2, \ldots, x_d).$$

Also, let $|x| = (\sum_{i=1}^d x_i^2)^{1/2}$ and $\omega(x) = (1/(2\pi)^{d/2})\exp(-\frac{1}{2}|x|^2)$. Define

$$L_\omega^p(\Lambda^d) = \{v \mid v \text{ is measurable and } \|v\|_{L_\omega^p(\Lambda^d)} < \infty\}.$$

Let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_d)$ be a multi-index and

$$\partial_x^{\boldsymbol{\alpha}} v(x) = \frac{\partial^{|\boldsymbol{\alpha}|} v}{\partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d}}(x),$$

where $|\,\boldsymbol{\alpha}\,| = \sum_{j=1}^d \alpha_j$. For any nonnegative integer r,

$$H_\omega^r(\Lambda^d) = \{v \mid \partial_x^{\boldsymbol{\alpha}} v \in L_\omega^2(\Lambda^d),\ 0 \le |\boldsymbol{\alpha}| \le r\}.$$

The semi-norm $|v|_{H_\omega^r(\Lambda^d)}$ and the norm $\|v\|_{H_\omega^r(\Lambda^d)}$ of $H_\omega^r(\Lambda^d)$ are the natural extensions of the one-dimensional case (see [10] and [16] and references therein).

First, we deal with the one-dimensional case. To this end, we need a lemma.

Lemma 3. *For any positive integer r, if $v \in H_\omega^r(\Lambda)$ then, for sufficiently large K,*

$$\|v - P_K v\|_{L_\omega^2(\Lambda)} \le \frac{1}{\sqrt{(K+1)K\cdots(K-r+2)}}\|v\|_{H_\omega^r(\Lambda)}.$$

Proof. We note that, by the Plancherel theorem,

$$\|v - P_K v\|_{L_\omega^2(\Lambda)}^2 = \sum_{l=K+1}^{\infty} a_l^2,$$

where $a_l = \int_{\mathbb{R}} v(x)\psi_l(x)\omega(x)\,\mathrm{d}x$, $\omega(x) = (1/\sqrt{2\pi})\exp(-\frac{1}{2}x^2)$. Now, by (1), (2), and integration by parts,

$$\begin{aligned}
\int_{\mathbb{R}} v(x)\psi_l(x)\omega(x)\,\mathrm{d}x &= \int_{\mathbb{R}} v(x)\bigg[\frac{x}{\sqrt{l}}\psi_{l-1}(x) - \frac{\sqrt{l-1}}{\sqrt{l}}\psi_{l-2}(x)\bigg]\omega(x)\,\mathrm{d}x \\
&= \frac{1}{\sqrt{l}}\int_{\mathbb{R}} xv(x)\psi_{l-1}(x)\omega(x)\,\mathrm{d}x - \frac{1}{\sqrt{l}}\int_{\mathbb{R}} v(x)\psi_{l-1}'(x)\omega(x)\,\mathrm{d}x \\
&= \frac{1}{\sqrt{l}}\int_{\mathbb{R}} v'(x)\psi_{l-1}(x)\omega(x)\,\mathrm{d}x.
\end{aligned}$$

By repeating the calculation, we have

$$\int_{\mathbb{R}} v(x)\psi_l(x)\omega(x)\,\mathrm{d}x = \frac{1}{\sqrt{l}\sqrt{l-1}\cdots\sqrt{l-r+1}}\int_{\mathbb{R}} v^{(r)}(x)\psi_{l-r}(x)\omega(x)\,\mathrm{d}x.$$

Thus,

$$\bigg|\int_{\mathbb{R}} v(x)\psi_l(x)\omega(x)\,\mathrm{d}x\bigg| \le \frac{1}{\sqrt{(K+1)K\cdots(K-r+2)}}\bigg|\int_{\mathbb{R}} v^{(r)}(x)\psi_{l-r}(x)\omega(x)\,\mathrm{d}x.\bigg|$$

Therefore,

$$\|v - P_K v\|_{L_\omega^2(\Lambda)}^2 \le \frac{1}{\sqrt{(K+1)K\cdots(K-r+2)}}\|v^{(r)}\|_{L^2_{\omega(\Lambda)}}^2,$$

which completes the proof. $\square$

If C_n is in $H^r_\omega(\Lambda)$ for each r, the error $\mathbb{E}[(C_{K,n}-C_n)^2]$ converges faster than any polynomial order and we may expect exponential decay of the error in L^2. One can find the same result for $\omega(x)=\mathrm{e}^{-x^2}$ in [16]. However, the proof there is not applicable here.

We now address a convergence rate for the case where $d=1$.

Proposition 2. *Suppose that $K=((1-\delta)/c)\log N$. Then, if $C_n\in H^r_\omega(\Lambda)$ for some positive integer r and each $n=1,\dots,m-1$, then the algorithm converges at least as fast as $O((\log N)^{-r/2})$ in L^2.*

Proof. By the proof of Theorem 2,

$$\mathbb{E}[(\hat{C}_{N,K,n}(\xi_n)-C_n(\xi_n))^2]\le C_1\frac{K^{2m}3^{2mK}}{N}+C_2\sum_{l=2}^{m-n}\mathbb{E}[(C_{K,m-l}(\xi_{m-l})-C_{m-l}(\xi_{m-l}))^2].$$

Then, using (1), (2), and Lemma 3 completes the proof. □

We note this proposition allows us to choose N and, in turn, K so as to obtain an approximation within a given error, which is convenient in practice.

We now consider the multi-dimensional case. To this end, we add one more condition to our multi-index scheme, the graded lexicographic order; given an expansion order L, we use a truncated basis $\{\boldsymbol{\psi}_k\colon|\alpha(k)|\le L\}$. With this new scheme, we thus have $\sum_{k=0}^{K}a_k\boldsymbol{\psi}_k$, where $1+K=(d+L)!/d!L!$ for $L=0,1,2,\dots$.

Lemma 4. *For any positive integer r, if $v\in H^r_\omega(\Lambda^d)$, for sufficiently large K,*

$$\|v-P_Kv\|_{L^2_\omega(\Lambda^d)}\le\frac{1}{\sqrt{(L/d)(L/d-1)\cdots(L/d-(r-1))}}\|v\|_{H^r_\omega(\Lambda^d)}.$$

Proof. We note, by the Plancherel theorem,

$$\|v-P_Lv\|^2_{L^2_\omega(\Lambda)}=\sum_{|\alpha(k)|>L}a^2_{\alpha(k)},$$

where $a_{\alpha(k)}=\int_{\mathbb{R}^d}v(x)\psi_{\alpha(k)_1}(x_1)\cdots\psi_{\alpha(k)_d}(x_d)\omega(x)\,\mathrm{d}x$. For $|\alpha(k)|=L+1$, there exists at least one component $\alpha(k)_i$ such that $\alpha(k)_i\ge|\alpha(k)|/d$. Suppose that, for some $k,\alpha(k)_1\ge(L+1)/d$. Then, by Lemma 3, we have

$$\begin{aligned}&\int_{\mathbb{R}^d}v(x)\psi_{\alpha(k)_1}(x_1)\cdots\psi_{\alpha(k)_d}(x_d)\omega(x)\,\mathrm{d}x\\&\quad\le\left(\sqrt{\left(\frac{L+1}{d}\right)\left(\frac{L+1}{d}-1\right)\cdots\left(\frac{L+1}{d}-(r-1)\right)}\right)^{-1}\\&\qquad\times\int_{\mathbb{R}^d}\frac{\partial^{(r)}v(x)}{\partial x_1}\psi_{\alpha(k)_1}(x_1)\psi_{\alpha(k)_2}(x_2)\cdots\psi_{\alpha(k)_d}(x_d)\omega(x)\,\mathrm{d}x.\end{aligned}$$

Hence, we have

$$\|v-P_Lv\|^2_{L^2_\omega(\Lambda^d)}\le\frac{1}{(L/d)(L/d-1)\cdots(L/d-(r-1))}\sum_{i=1}^{d}\left\|\frac{\partial^{(r)}v(x)}{\partial x_i}\right\|^2_{L^2_{\omega(\Lambda^d)}},$$

which completes the proof. □

We address a rate of convergence of the algorithm for the multi-dimensional case.

https://doi.org/10.1017/jpr.2017.9 Published online by Cambridge University Press

Proposition 3. *Suppose that $L = ((1-\delta)/c)\log N$. Then, if $C_n \in H^r_\omega(\Lambda^d)$ for some positive integer r and each $n = 1, \dots, m-1$, then the algorithm converges at least as fast as $O((\log N)^{-r/2})$ in L^2.*

Proof. By the proof of Theorem 2,

$$\begin{aligned}\mathbb{E}[(\hat{C}_{N,K,n}(\boldsymbol{\xi}_n) - C_n(\boldsymbol{\xi}_n))^2] \\ \le C_1 \frac{((d+L)!/d!\,L!)^{2m} 3^{2mL}}{N} + C_2 \sum_{l=2}^{m-n} \mathbb{E}[(C_{K,m-l}(\boldsymbol{\xi}_{m-l}) - C_{m-l}(\boldsymbol{\xi}_{m-l}))^2].\end{aligned}$$

Then, using (1), (2), and Lemma 4 completes the proof. □

6. Concluding remarks

In this paper we improved on the results of [9]. First of all, we proved the L^2-convergence of the quasi-regression Monte Carlo method to the true price of an American option under the setting where the number of paths and number of basis functions increase together and the dimension of the underlying assets is greater than one. Secondly, we have shown that the highest possible number of basis functions for N paths is $O(\log N)$ to achieve convergence in implementing the method even under multiple undelyings. Finally, we proposed a rate of convergence considering the regularity of the continuation value function.

For further research, one question is to find a sharper convergence rate than the one proposed in this paper. It amounts to finding a sharper bound on the error between the finite truncation of the continuation value function and itself. It could be a challenging problem. Another question is an extension of the results in [7] in the same manner as here. The critical hardness could be how to obtain L^p-asymtotics on the basis function used in [7], see Appendix B in that paper.

Appendix A.

The idea behind the proofs of Lemmas 1 and 2 for Theorem 1 is similar to that for the case of BM in [9], although the details are different because of differences in the number of underlyings and the assumptions on the fourth moments that are less restrictive than in [9].

Proof of Lemma 1. In this proof we drop the boldface notation for convenience.

(i) It is obvious since $\hat{V}_m(S^i_m) = h_m(S^i_m) = \bar{V}_m(S^i_m)$.

(ii) By the Cauchy–Schwarz inequality,

$$\begin{aligned}(\hat{\beta}_{n,k} - \tilde{\beta}_{n,k})^2 &= \bigg(\frac{1}{N}\sum_{i=1}^N \hat{V}_{n+1}(S^i_{n+1})\psi_k(\xi^i_n) - \frac{1}{N}\sum_{i=1}^N \bar{V}_{n+1}(S^i_{n+1})\psi_k(\xi^i_n)\bigg)^2 \\ &\le \frac{1}{N}\sum_{i=1}^N \psi_k^2(\xi^i_n)(\hat{V}_{n+1}(S^i_{n+1}) - \bar{V}_{n+1}(S^i_{n+1}))^2.\end{aligned}$$

Then, by noting that

$$\begin{aligned}&|\hat{V}_{n+1}(S^i_{n+1}) - \bar{V}_{n+1}(S^i_{n+1})| \\ &\quad = |\max\{h_{n+1}(S^i_{n+1}), \hat{C}_{N,K,n+1}(S^i_{n+1})\} - \max\{h_{n+1}(S^i_{n+1}), \bar{C}_{N,K,n+1}(S^i_{n+1})\}| \\ &\quad \le ||C_{N,K,n+1}(S^i_{n+1}) - \bar{C}_{N,K,n+1}(S^i_{n+1})|,\end{aligned}$$

and

$$(\hat{C}_{N,K,n+1}(S^i_{n+1}) - \bar{C}_{N,K,n+1}(S^i_{n+1}))^2 = \biggl(\sum_{k=0}^{K}(\hat{\beta}_{n+1,k} - \bar{\beta}_{n+1,k})\psi_k(\xi^i_{n+1})\biggr)^2$$
$$\leq (K+1)\sum_{k=0}^{K}(\hat{\beta}_{n+1,k} - \bar{\beta}_{n+1,k})^2\psi_k^2(\xi^i_{n+1}),$$

we have

$$\begin{aligned}\mathbb{E}[(\hat{\beta}_{n,k} - \tilde{\beta}_{n,k})^2] &\leq (K+1)\mathbb{E}\biggl[\sum_{l=0}^{K}\psi_k^2(\xi^i_n)\psi_l^2(\xi^i_{n+1})(\hat{\beta}_{n+1,l} - \bar{\beta}_{n+1,l})^2\biggr] \\ &= (K+1)\sum_{l=0}^{K}\mathbb{E}[\psi_k^2(\xi^i_n)\psi_l^2(\xi^i_{n+1})]\mathbb{E}[(\hat{\beta}_{n+1,l} - \bar{\beta}_{n+1,l})^2] \\ &\leq (K+1)\sum_{l=0}^{K}\sqrt{\mathbb{E}[\psi_k^4(\xi^i_n)]}\sqrt{\mathbb{E}[\psi_k^4(\xi^i_{n+1})]}\mathbb{E}[(\hat{\beta}_{n+1,l} - \bar{\beta}_{n+1,l})^2] \\ &\leq (K+1)\max_{0\leq k\leq K}\mathbb{E}[\psi_k^4(\xi)]\sum_{l=0}^{K}\mathbb{E}[(\hat{\beta}_{n+1,l} - \bar{\beta}_{n+1,l})^2].\end{aligned}$$

Thus, letting $B_K = (K+1)\max_{0\leq k\leq K}\mathbb{E}[\psi_k^4(\xi)]$, we have

$$\mathbb{E}\biggl[\sum_{k=0}^{K}(\hat{\beta}_{n,k} - \tilde{\beta}_{n,k})^2\biggr] \leq B_K(K+1)\sum_{k=0}^{K}\mathbb{E}[(\hat{\beta}_{n+1,k} - \bar{\beta}_{n+1,k})^2].$$

Let $A_K = (K+1)^2\max_{0\leq k\leq K}\mathbb{E}[\psi_k^4(\xi)]$. Then, since

$$\mathbb{E}[(\hat{\beta}_{n+1,k} - \bar{\beta}_{n+1,k})^2] \leq 2\mathbb{E}[(\hat{\beta}_{n+1,k} - \tilde{\beta}_{n+1,k})^2] + 2\mathbb{E}[(\tilde{\beta}_{n+1,k} - \bar{\beta}_{n+1,k})^2],$$

we have

$$\mathbb{E}\biggl[\sum_{k=0}^{K}(\hat{\beta}_{n,k} - \tilde{\beta}_{n,k})^2\biggr] \leq 2A_K\sum_{k=0}^{K}\mathbb{E}[(\hat{\beta}_{n+1,k} - \tilde{\beta}_{n+1,k})^2] + 2A_K\sum_{k=0}^{K}\mathbb{E}[(\tilde{\beta}_{n+1,k} - \bar{\beta}_{n+1,k})^2].$$

By repeating the procedure, we reach

$$\begin{aligned}\mathbb{E}\biggl[\sum_{k=0}^{K}(\hat{\beta}_{n,k} - \tilde{\beta}_{n,k})^2\biggr] &\leq (2A_K)^{m-n-1}\sum_{k=0}^{K}\mathbb{E}[(\hat{\beta}_{m-1,k} - \tilde{\beta}_{m-1,k})^2] \\ &\quad + \sum_{l=1}^{m-n-1}(2A_K)^{m-n-l}\sum_{k=0}^{K}\mathbb{E}[(\tilde{\beta}_{m-l,k} - \bar{\beta}_{m-l,k})^2].\end{aligned}$$

Therefore , by (i), we finally have

$$\mathbb{E}\biggl[\sum_{k=0}^{K}(\hat{\beta}_{n,k} - \tilde{\beta}_{n,k})^2\biggr] \leq 2^{m-n-1}A_K^{m-n-1}\sum_{l=1}^{m-n-1}\sum_{k=0}^{K}\mathbb{E}[(\tilde{\beta}_{m-l,k} - \bar{\beta}_{m-l,k})^2],$$

which completes the proof. □

Proof of Lemma 2. For each $n \in \{1, \dots, m-1\}$,

$$\sum_{k=0}^{K} \mathbb{E}[(\tilde{\beta}_{n,k} - \bar{\beta}_{n,k})^2] \le C \frac{(K+1)^3}{N} \Big(\max_{0 \le k \le K} \sqrt{\mathbb{E}[\boldsymbol{\psi}_k^4]} + \max_{0 \le k \le K} \mathbb{E}[\boldsymbol{\psi}_k^4] \Big).$$

Note that $\mathbb{E}[\tilde{\beta}_{n,k}] = \bar{\beta}_{n,k}$. Then, for $n \in \{1, \dots, m-2\}$,

$$\begin{aligned}
\sum_{k=0}^{K} \mathbb{E}[(\tilde{\beta}_{n,k} - \bar{\beta}_{n,k})^2] &= \sum_{k=0}^{K} \frac{1}{N} \operatorname{var}(\psi_k(\xi_n) \bar{V}_{n+1}(S_{n+1})) \\
&\le \sum_{k=0}^{K} \frac{1}{N} \mathbb{E}[\psi_k^2(\xi_n) \bar{V}_{n+1}^2(S_{n+1})] \\
&\le \sum_{k=0}^{K} \frac{1}{N} \mathbb{E}[\psi_k^2(\xi_n) \max\{h_{n+1}^2(S_{n+1}), \bar{C}_{n+1}^2(S_{n+1})\}] \\
&\le \sum_{k=0}^{K} \frac{1}{N} \mathbb{E}[\psi_k^2(\xi_n)(h_{n+1}^2(S_{n+1}) + \bar{C}_{n+1}^2(S_{n+1}))] \\
&= \frac{1}{N} \sum_{k=0}^{K} \mathbb{E}[\psi_k^2(\xi_n) h_{n+1}^2(S_{n+1})] + \frac{1}{N} \sum_{k=0}^{K} \mathbb{E}[\psi_k^2(\xi_n) \bar{C}_{n+1}^2(S_{n+1})].
\end{aligned}$$

Thus, by the Cauchy–Schwarz inequality,

$$\begin{aligned}
\sum_{k=0}^{K} \mathbb{E}[(\tilde{\beta}_{n,k} - \bar{\beta}_{n,k})^2] &\le \frac{1}{N} \sum_{k=0}^{K} \sqrt{\mathbb{E}[\psi_k^4(\xi)]} \sqrt{E[h_{n+1}^4(S_{n+1})]} \\
&\quad + \frac{1}{N} \sum_{k=0}^{K} \mathbb{E}\bigg[\psi_k^2(\xi_n) \bigg(\sum_{l=0}^{K} \bar{\beta}_{n+1,l} \psi_l(\xi_{n+1}) \bigg)^2 \bigg] \\
&\le \frac{1}{N}(K+1) \max_{1 \le n \le m} \sqrt{\mathbb{E}[h_n^4(S_n)]} \max_{0 \le k \le K} \sqrt{\mathbb{E}[\psi_k^4(\xi)]} \\
&\quad + \frac{1}{N}(K+1) \sum_{k=0}^{K} \mathbb{E}\bigg[\psi_k^2(\xi_n) \bigg(\sum_{l=0}^{K} \bar{\beta}_{n+1,l}^2 \psi_l^2(\xi_{n+1}) \bigg) \bigg].
\end{aligned}$$

Then, by noting that $\sum_{k=0}^{K} \bar{\beta}_{n+1,k}^2 = \mathbb{E}[(P_K \bar{V}_{n+2})^2] \le \mathbb{E}[\bar{V}_{n+2}^2]$ and invoking Assumption 2,

$$\begin{aligned}
\sum_{k=0}^{K} \mathbb{E}[(\tilde{\beta}_{n,k} - \bar{\beta}_{n,k})^2] &\le \frac{1}{N}(K+1) \max_{1 \le n \le m} \sqrt{\mathbb{E}[h_n^4(S_n)]} \max_{0 \le k \le K} \sqrt{\mathbb{E}[\psi_k^4(\xi)]} \\
&\quad + \frac{1}{N} \|\bar{V}_{n+2}\|_{L^2}^2 (K+1) \sum_{k=0}^{K} \mathbb{E}\bigg[\psi_k^2(\xi_n) \bigg(\sum_{l=0}^{K} \psi_l^2(\xi_{n+1}) \bigg) \bigg] \\
&\le \frac{1}{N}(K+1) \max_{1 \le n \le m} \sqrt{\mathbb{E}[h_n^4(S_n)]} \max_{0 \le k \le K} \sqrt{\mathbb{E}[\psi_k^4(\xi)]} \\
&\quad + \frac{1}{N} \mathbb{E}[\bar{V}_{n+1}^2](K+1)^3 \max_{0 \le k \le K} \mathbb{E}[\psi_k^4(\xi)] \\
&\le C(K+1)^3 \frac{1}{N} \Big(\max_{0 \le k \le K} \sqrt{\mathbb{E}[\psi_k^4(\xi)]} + \max_{0 \le k \le K} \mathbb{E}[\psi_k^4(\xi)] \Big)
\end{aligned}$$

Table 1.

$\lvert \boldsymbol{i} \rvert$	Multi-index $\boldsymbol{i}$	Single index k
1	(1 0 0 0)	2
	(0 1 0 0)	3
	(0 0 1 0)	4
	(0 0 0 1)	5
2	(2 0 0 0)	6
	(1 1 0 0)	7
	(1 0 1 0)	8
	(1 0 0 1)	9
	(0 2 0 0)	10
	(0 1 1 0)	11
	(0 1 0 1)	12
	(0 0 2 0)	13
	(0 0 1 1)	14
	(0 0 0 2)	15
3	(3 0 0 0)	16
	(2 1 0 0)	17
	(2 0 1 0)	18
	$\vdots$	$\vdots$

for some constant $C > 0$. Since $\max\{h_m^2(S_m), \bar{C}_m^2(S_m)\} = h_m^2(S_m)$, the estimate also holds for $n = m - 1$, which completes the proof. $\square$

An example of graded lexicographic ordering. In Table 1 we present a graded lexicographic ordering of the multi-index $\boldsymbol{i}$ in $d = 4$ dimensions.

Acknowledgements

The author would like to thank Dr Sihun Jo at Woosuk University for his help during preparation of the paper and the anonymous referees for invaluable comments which lead to significant improvements in the paper.

References

[1] Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions*, 9th edn. Springer, New York.

[2] Carriere, J. F. (1996). Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance Math. Econom.* **19,** 19–30.

[3] Clément, E., Lamberton, D. and Protter, P. (2002). An analysis of a least squares regression method for American option pricing. *Finance Stoch.* **6,** 449–471.

[4] Egloff, D. (2005). Monte Carlo algorithms for optimal stopping and statistical learning. *Ann. Appl. Prob.* **15,** 1396–1432.

[5] Egloff, D., Kohler, M. and Todorovic, N. (2007). A dynamic look-ahead Monte Carlo algorithm for pricing Bermudan options. *Ann. Appl. Prob.* **17,** 1138–1171.

[6] Ernst, O. G., Mugler, H. I., Starkoff, H. J. and Ullmann, E. (2012). On the convergence of generalized polynomial chaos expansions. *ESIM Math. Modelling Numer. Anal.* **46,** 317–339.

[7] Gerhold, S. (2011). The Longstaff-Schwartz algorithm for Levy models: results on fast and slow convergence. *Ann. Appl. Prob.* **21,** 589–608.

[8] Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering*. Springer, New York.

[9] Glasserman, P. and Yu, B. (2004). Number of paths versus number of basis functions in American option pricing. *Ann. Appl. Prob.* **14,** 2090–2119.
[10] Guo, B.-Y. (1999). Error estimation of Hermite spectral method for nonlinear partial differential equations. *Math. Comput.* **68,** 1067–1078.
[11] Lars, L.-C. (2002). L^p-norms of Hermite polynomials and an extremal problem on Wiener chaos. *Ark. Mat.* **40,** 133–144.
[12] Longstaff, F. A. and Schwartz, E. S. (2001). Valuing American options by simulation: a simple least-squares approach. *Rev. Financial Studies* **14,** 113–147.
[13] Owen, A. B. (2000). Assessing linearity in high dimensions. *Ann. Statist.* **28,** 1–19.
[14] Stentoft, L. (2004). Convergence of the least squares Monte Carlo approach to American option valuation. *Manag. Sci.* **50,** 1193–1203.
[15] Tsitsiklis, J. N. and Van Roy, B. (2001). Regression methods for pricing complex American-style options. *IEEE Trans. Neural Networks* **12,** 694–703.
[16] Xu, C.-L. and Guo, B.-Y. (2003). Hermite spectral and pseudospectral methods for nonlinear partial differential equations. *Comput. Appl. Math.* **22,** 167–193.
[17] Zanger, D. Z. (2013). Quantitative error estimates for a least-squares Monte Carlo algorithm for American option pricing. *Finance Stoch.* **17,** 503–534.