CIRJE-F-1178

Deep Asymptotic Expansion: Application to Financial Mathematics

Yuga Iguchi
MUFG Bank
UCL London

Riu Naito
Japan Post Insurance
Hitotsubashi University

Yusuke Okano
SMBC Nikko Securities

Akihiko Takahashi
The University of Tokyo

Toshihiro Yamada
Hitotsubashi University
Japan Science and Technology Agency

November 2021; Revised in February 2022

CIRJE Discussion Papers can be downloaded without charge from:
http://www.cirje.e.u-tokyo.ac.jp/research/03research02dp.html

Discussion Papers are a series of manuscripts in their draft form. They are not intended for circulation or distribution except as indicated by the author. For that reason Discussion Papers may not be reproduced or distributed without the written consent of the author.
Deep Asymptotic Expansion: Application to Financial Mathematics

Yuga Iguchi
MUFG Bank
Tokyo, Japan
yuga.iguchi@gmail.com

Riu Naito
Japan Post Insurance
& Hitotsubashi University
Tokyo, Japan
riu.naito@gmail.com

Yusuke Okano
SMBC Nikko Securities
Tokyo, Japan
yusukensw@gmail.com

Akihiko Takahashi
University of Tokyo
Tokyo, Japan
akihiko@e.u-tokyo.ac.jp

Toshihiro Yamada
Hitotsubashi University
& JST
Tokyo, Japan
toshihiro.yamada@r.hit-u.ac.jp

Abstract—The paper proposes a new computational scheme for diffusion semigroups based on an asymptotic expansion with weak approximation and deep learning algorithm to solve high-dimensional Kolmogorov partial differential equations (PDEs). In particular, we give a spatial approximation for the solution of \( d \)-dimensional PDEs on a range \([a, b]^d\) without suffering from the curse of dimensionality.

Index Terms—Deep learning, Asymptotic expansion, Weak approximation, Kolmogorov PDEs, Malliavin calculus, Curse of dimensionality

I. INTRODUCTION

Kolmogorov partial differential equations (PDEs) are widely used in various fields such as physics, engineering and financial mathematics. In general there are no closed form solutions except for a few special cases. Hence, numerical methods are usually required to solve Kolmogorov PDEs.

As classical schemes for solving Kolmogorov PDEs, finite element and finite difference methods are well known. These spatial approximation schemes work only for lower (typically from 1 to 3) dimensions since the computational complexity grows exponentially in the dimension of target Kolmogorov PDEs. In other words, finite element/difference methods suffer from the curse of dimensionality.

Instead, Monte Carlo methods can be applied to high dimensional cases due to the advantage of overcoming the curse of dimensionality. In perspective of solving Kolmogorov PDEs, some discretization methods (weak and strong approximations) of stochastic differential equations are used with Monte Carlo methods. However, Monte Carlo method provides an approximation at a fixed single point for the solution, that is, it does not give a spatial approximation.

We also point out that there exist closed form approximations for solutions of Kolmogorov PDEs such as asymptotic expansion methods. In particular, an expectation of a diffusion process, satisfying a Kolmogorov PDE with weak approximation. We call this approximation the deep asymptotic expansion (Deep AE) for short.

The paper is organized as follows. In the next section, we introduce an asymptotic expansion approach for solving Kolmogorov PDEs with weak approximation. Section III describes a deep learning-algorithm for our asymptotic expansion method. Section IV shows numerical results to demonstrate the efficiency of the proposed method. Appendix provides proofs for propositions in the main text.

II. ASYMPTOTIC EXPANSION AND WEAK APPROXIMATION

On a filtered probability space \((\Omega, \mathcal{F}, \{\mathcal{F}_t\}_t, P)\), let \(W = \{W_t\}_{t \geq 0}\) be a \(d\)-dimensional \(\mathcal{F}_t\)-Brownian motion. For \(t \geq 0\)
0 and $T > t$, let $X_{t,x}^{t,x,ε}, s \in [t,T], x \in \mathbb{R}^d$ be the solution of

$$X_{t,x}^{t,x,ε} = x + \int_t^s \mu(r, X_{t,x}^{t,x,ε})dr + \varepsilon \sum_{i=1}^d \sigma_i(r, X_{t,x}^{t,x,ε})dW_i,$$

where $\varepsilon \in (0,1]$ and $\mu : [0,T] \times \mathbb{R}^d \to \mathbb{R}, \sigma_i : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d, i = 1, \ldots, d$ are continuous and bounded in $t$ and continuously differentiable in $x$ with bounded derivatives of any order. Let $\{P_{t,x}^{ε}\}_{t,x} \geq t$ be a two-parameter semigroup of linear operators given by

$$(P_{t,x}^{ε}f)(x) = E[f(X_{t,x}^{t,x,ε})], \quad s \geq t, \quad x \in \mathbb{R}^d,$$  

for a continuous function $f : \mathbb{R}^d \to \mathbb{R}$. The aim of this paper is to show an approximation scheme for the function $x \mapsto (P_{t,x}^{ε}f)(x)$ where $f : \mathbb{R}^d \to \mathbb{R}$ is a continuous function. The $d$-dimensional process $X_{t,x}^{t,x,ε} = (X_{t,x}^{t,x,ε,1}, \ldots, X_{t,x}^{t,x,ε,d})$ can be expanded as follows: for $i = 1, \ldots, d$,

$$X_{t,x}^{t,x,ε,i} = X_{t,x}^{t,x,0,i} + \varepsilon X_{t,x}^{t,x,1,i} + \varepsilon^2 X_{t,x}^{t,x,2,i} + \ldots$$

in Malliavin sense, where $X_{t,x}^{t,x,0,i}$ is the solution of $X_{t,x}^{t,x,0,i} = x + \int_t^s \mu_i(r, X_{t,x}^{t,x,0,i})dr$, and $X_{t,x}^{t,x,i}, k \in \mathbb{N}$ given by $X_{t,x}^{t,x,i} = \mathbb{E}[\partial x X_{t,x}^{t,x,i}], \quad i = 0, \ldots, k$.

Let us define $X_{t,x}^{t,x,ε} = X_{t,x}^{t,x,0} + \varepsilon X_{t,x}^{t,x,1} + \varepsilon^2 X_{t,x}^{t,x,2} + \ldots$ in the following Wiener integrals: $X_{t,x}^{t,x,i} = \sum_{n=1}^i \int_0^s J_{t,x}^{n,0}(J_{t,x}^{n,0})^{-1}dW_i$, with $J_{t,x}^{n,0} = \partial x X_{t,x}^{t,x,n}, r \geq t$.

We introduce an expansion of $P_{t,T}^{ε}f$ with respect to the parameter $ε$ and then give a second-order discretization with respect to the number of time-steps $n$. We only use polynomials of the Gaussian random variable $X_{t,x}^{t,x,i}$ up to the third order on each subinterval $[t_i, t_{i+1}], i = 0, 1, \ldots, n-1$, where $t_i = t + i(T-t)/n, i = 0, 1, \ldots, n$ are the time-grid of the uniform partition on $[t,T]$. Let $\{Q_{t_i,x}^{ε}\}_{t_i} \geq t$ be linear operators given by

$$(Q_{t_i,x}^{ε}f)(x) = E[f(X_{t,x}^{t,x,ε})(W_{t,x}^{t,x,ε})], \quad s \geq t, \quad x \in \mathbb{R}^d,$$  

for a continuous and bounded function $f : \mathbb{R}^d \to \mathbb{R}$, where $W_{t,x}^{t,x,ε}$ is a Malliavin weight given in Appendix.

**Theorem 1.** Then, there exists $C > 0$ such that

$$\left\| P_{t,T}^{ε}f - Q_{t_1,x_1}^{ε} \cdots Q_{t_n,x_n}^{ε}f \right\|_\infty \leq \varepsilon^2 C \|f\|_\infty \frac{1}{n^2},$$

for any $ε > 0, n \geq 1$ and continuous bounded function $f : \mathbb{R}^d \to \mathbb{R}$.

**Proof of Theorem 1.** See Section V □

Here, the approximation can be expressed as

$$(Q_{t_i,x}^{ε} \cdots Q_{t_n,x}^{ε}f)(x) = E[f(\bar{X}_{t_i,x}^{t,x,ε}(n)) \prod_{i=1}^n W_{t_i}^{t_i-1,\bar{X}_{t_i,x}^{t,x,ε}(n)}, x \in \mathbb{R}^d, \quad \bar{X}_{t_i,x}^{t,x,ε}(n) = \bar{X}_{t_i,x}^{t_i-1,\bar{X}_{t_i,x}^{t,x,ε}(n)}, i = 1, \ldots, n.$$

Solutions of Kolmogorov PDEs are approximated in the following way. Let $u^ε : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ be a function given by $u^ε(t,x) = E[f(X_{t,x}^{t,x,ε}(n))]$ with a continuous function $f : \mathbb{R}^d \to \mathbb{R}$ of polynomial growth order, which satisfies

$$(\partial_t + L_t^{ε})u^ε(t,x) = 0, \quad u^ε(T,x) = f(x),$$

where

$$L_t^{ε} = \sum_{j=1}^d \mu_j(t,\cdot) \frac{d}{dx_j} + \frac{ε^2}{2} \sum_{i,j=1}^d \sigma_i(t,\cdot) \sigma_j(t,\cdot) \frac{d^2}{dx_i dx_j}.$$  

Then, there exist $C > 0$ and $q > 0$ such that

$$|u^ε(t,x) - E[f(X_{t,x}^{t,x,ε}(n)) \prod_{i=1}^n W_{t_i}^{t_i-1,\bar{X}_{t_i,x}^{t,x,ε}(n)}]| \leq ε^2 C(1 + |x|^q) \frac{1}{n^2},$$

for any $ε > 0, n \geq 1$ and $x \in \mathbb{R}^d$.

**III. DEEP LEARNING-BASED APPROXIMATION**

Let $a \in \mathbb{R}, b \in (a, \infty), t > 0, T > t, n \in \mathbb{N}$ and $ξ : [a, b]^d \to [a, b]^d$ be measurable uniformly distributed random variable. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a continuous function with polynomial growth. We define $X_{t,x}^{ξ}(n) = X_{t,x}^{t,x,ξ(0)}(n)$. Then, the following holds:

$$v^* = \arg\min_{v \in C([a,b]^d)} E \left[ \left| v(ξ) - E[f(X_{t,x}^{ξ}(n)) \prod_{i=1}^n W_{t_i}^{t_i-1,\bar{X}_{t_i,x}^{ξ(0)}(n)}]\right|^2 \right],$$

and it holds that for all $x \in [a, b]^d$,

$$v^*(x) = Q_{t_1,x_1} \cdots Q_{t_n,x_n} f(x).$$

With the above representation, the function $P_{t,T}^{ε}f$ can be approximated using deep learning. Let $L^r \ni x \mapsto L_r(x) \in \mathbb{R}^r$ be the Rectified Linear Unit (ReLU) activation function given by

$$L_r(x) = \left( \begin{array}{c} \theta_{q+1} & \cdots & \theta_{q+p} \\ \vdots & \ddots & \vdots \\ \theta_{q+(p-1)p+1} & \cdots & \theta_{q+lp} \end{array} \right) \left( \begin{array}{c} x_1 \\ \vdots \\ x_p \end{array} \right) + \left( \begin{array}{c} \theta_{q+lp+1} \\ \vdots \\ \theta_{q+lp+\ell} \end{array} \right).$$

Let $s \in \{3, 4, 5, \ldots\}$ such that $\sum_{k=1}^s d_k(d_k-1) \leq \nu$ for $d_0 = d, \quad d_1 = 1, \quad d_2 = 1, \ldots, d_{s-1} = \mathbb{N}$, and then we have

$$P_{t,T}^{ε}f \approx Q_{t_1,T}^{ε[n]} θ^{ε[n]} f,$$  

where $Q_{t_1,T}^{ε[n]} θ^{ε[n]} f$ is given by

$$Q_{t_1,T}^{ε[n]} θ^{ε[n]} f = (A_{d_k-1,d_k}^{θ_{q+1},\sum_{k=1}^s d_k(d_k-1)} \circ L_{d_k-1} \circ A_{d_k-2,d_k-1}^{θ_{q+2},\sum_{k=1}^s d_k(d_k-1)} \circ \cdots \circ L_{d_k} \circ A_{d_k,0}^{θ_{q+p},\sum_{k=1}^s d_k(d_k-1)}\circ L_{d_k} \circ A_{d_k,0}^{θ_{q+p+1},\sum_{k=1}^s d_k(d_k-1)})(x), \quad x \in \mathbb{R}^d$$

with $θ^{ε[n]}$ satisfying

$$θ^{ε[n]} = \arg\min_{θ^{ε[n]}} E \left[ \left| Q_{t_1,T}^{ε[n]} θ^{ε[n]} f(ξ) - f(X_{t,x}^{ξ}(n)) \prod_{i=1}^n W_{t_i}^{t_i-1,\bar{X}_{t_i,x}^{ξ(0)}(n)}\right|^2 \right].$$

The function $Q_{t_1,T}^{ε[n]} θ^{ε[n]} f : \mathbb{R}^d \to \mathbb{R}$ with $θ^{ε[n]}$ represents an artificial neural network with $s+1$ layers (1 input layer with $d$ neurons, $k$-th hidden layers with $d_k$ neurons for each $k = 1, \ldots, s-1$, and 1 output layer with 1 neuron).
IV. NUMERICAL EXAMPLES FOR FINANCIAL MATHEMATICS

In the section, we apply the proposed method to the following $d$-dimensional Kolmogorov PDE:

$$\left(\partial_t + L_t^a\right)u^x(t,x) = \phi, \quad u^x(T,x) = f(x),$$

where $L_t^a$ is a second order differential operator given by

$$L_t^a \varphi(x) = \sum_{j=1}^{d} \partial_j^2 \varphi(x) + \sum_{i,j=1}^{d} \sigma_{ij}^2 x_i x_j \frac{\partial^2 \varphi(x)}{\partial x_i \partial x_j}$$

(12)

for a smooth function $\varphi$ and $x \in \mathbb{R}^d$, and $f$ is a continuous function which is specified in the following subsections.

A. Numerical error of Deep AE

First, we evaluate the error for an entire region $[a,b]^d$. Let $d = 10$, $a = 99.0$, $b = 101.0$, $r = 0.01$, $\delta_t^2 = \delta_t^4$, $\varepsilon = 0.2$, $T = 5.0$, and $f_4 : \mathbb{R}^d \rightarrow \mathbb{R}$ be a function $y \mapsto \max\{\max\{y_1 - K, 0\}, \ldots, \max\{y_d - K, 0\}\}$ with $K = 100.0$, where $\delta_t^6$ is the Kronecker delta. As an example, we apply the proposed second order asymptotic expansion with second order scheme (Deep AE) and continuous uniformly distributed random variable $\xi : \Omega \rightarrow [a,b]^d$.

In Table II, numerical errors and runtimes of Deep EM of Beck et al. (2018) are compared with those of the current scheme. As an error analysis after the solution of Kolmogorov PDE is estimated by each scheme, we compute $\max_{x \in \{y_0, \ldots, y_k\}} \frac{\text{Ref}(x) - \text{Deep AE}(x)}{\text{Ref}(x)}$ and $\max_{x \in \{y_0, \ldots, y_k\}} \frac{\text{Ref}(x) - \text{Deep EM}(x)}{\text{Ref}(x)}$, where $y_l = (a + (b-a)i/k, \ldots, a + (b-a)i/k) \in \mathbb{R}^d$, $k = 20$, $i < k$, $\text{Deep AE}(x)$ and $\text{Deep EM}(x)$ represent numerical values of Deep AE and Deep EM, respectively, and $\text{Ref}(x)$, $x \in \{y_0, \ldots, y_k\}$ are computed by Monte Carlo simulations with the number of paths $10^8$ and the explicit solution of $X^x$ obtained by Itô formula. The table below shows that the convergence of our scheme is faster than that of deep EM as spatial approximation.

B. Weak convergence

Next, we check the rate of weak convergence based on the theoretical estimate (6). As in the previous subsection, we compare the accuracies of the proposed scheme with those of Deep EM of Beck et al. (2018). In the experiments, we first estimate the function $Q_{\xi_1^0} \circ \cdots \circ Q_{\xi_{n-1}^0} f$ on a region $[a,b]^d$ with continuous uniformly distributed random variable $\xi : \Omega \rightarrow [a,b]^d$, and compute $Q_{\xi_n^0} \circ \cdots \circ Q_{\xi_1^0} f(x)$ at $x \in [a,b]^d$. Then, we check the numerical error, where our reference value is computed by a Monte Carlo simulation with the number of paths $10^8$.

Figure 1 shows the result for $d = 10$, $a = 99.0$, $b = 101.0$, $r = 0.015$, $\sigma_t^2 = \delta_t^4$, $\varepsilon = 0.3$, $T = 2.0$, $f : \mathbb{R}^d \rightarrow \mathbb{R}$ given by $y \mapsto \max\{\max\{y_1 - K, 0\}, \ldots, \max\{y_d - K, 0\}\}$ with $K = 100.0$, and $x = (100, 0, \ldots, 100) \in [a,b]^d$, where the relative errors are plotted. Here, we use 1 input layer, 2 hidden layers, 1 output layer with neurons $(d, d + 50, d + 50, 1)$ in the deep learning computation. Also, the batch size $M$, the train steps $J$ and learning rate $\gamma_j$, $j \leq J$ in the stochastic gradient descent method are taken as $M = 8192$, $J = 50000$ and $\gamma_j = 10^{-2} 1_{[0, 0.2]}(j) + 10^{-3} 1_{[0.2, 0.6]}(j) + 10^{-4} 1_{[0.6, 1]}(j)$, $j = 0, 1, \ldots, J$. In Table II, numerical errors and runtimes of the schemes is shown.

TABLE I

<table>
<thead>
<tr>
<th>Number of train steps</th>
<th>Error for Deep AE ($n = 2^k$)</th>
<th>Runtime for Deep AE ($n = 2^k$)</th>
<th>Error for Deep EM ($n = 2^k$)</th>
<th>Runtime (s) for Deep EM ($n = 2^k$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500000</td>
<td>0.000609</td>
<td>495.54s</td>
<td>0.00671</td>
<td>20390.21s</td>
</tr>
</tbody>
</table>

Fig. 1. Weak convergence ($d = 10$)
Figure 2 shows the example for $d = 100, a = 99.0, b = 101.0, r = 0.015, \sigma^2 = \delta^2, \varepsilon = 0.2, T = 0.5, f : \mathbb{R}^d \rightarrow \mathbb{R}$ given by $y \mapsto \max\{\max\{y_1 - K, 0\}, \ldots, \max\{y_d - K, 0\}\}$ with $K = 100.0$, and $x = (100.0, \ldots, 100.0) \in [a, b]^d$, where the relative errors are plotted. Here, we use 1 input layer, 2 hidden layers, 1 output layer with neurons $(d, d + 50, d + 50, 1)$ in the deep learning computation. In the stochastic gradient descent method, the batch size $M$, the train steps $J$ and learning rate $\gamma(j)$, $j \leq J$ are taken as $M = 1024, J = 25000$ and $\gamma(j) = 5 \times 10^{-2} 1_{[0.02, 0.2)}(j) + 5 \times 10^{-5} 1_{(0.2, 0.6)}(j) + 5 \times 10^{-1} 1_{(0.6, 1]}(j)$, $j = 0, 1, \ldots, J$. Table II shows numerical errors and runtimes of the schemes.

Those figures and tables demonstrate that our Deep AE gives more accurate approximations than Deep EM and provides high performance in terms of runtime to achieve the same level of accuracies.

Throughout the numerical experiments, we have checked that the proposed scheme works as a spacial approximation in high-dimensional PDE models and the numerical results are consistent with theoretical parts given in Section II and III.

V. PROOF OF THEOREM 1

We prepare some notations on Malliavin calculus. Let $\mathbb{D}^{\infty}$ be the space of smooth Wiener functionals $F : C([0, T]; \mathbb{R}^d) \rightarrow \mathbb{R}$ in the sense of Malliavin. For a nondegenerate $F \in \mathbb{D}^{\infty}$, $G \in \mathbb{D}^{\infty}$ and a multi-index $\gamma$, there exists $H_\gamma(F, G) \in \mathbb{D}^{\infty}$ such that

\[ (IBP) \quad E[\partial^\gamma \varphi(F)G] = E[\varphi(F)H_\gamma(F, G)] \quad (13) \]


The following lemma is useful for the proof of Theorem 1.

**Lemma 1.** Let $0 \leq t < s$, $k \geq 3$, $\Delta_{t,s} := \{(t_1, \ldots, t_k) \in \mathbb{R}^k; t \leq t_1 < \cdots < t_k \leq s\}$ and $\alpha \in \{0, 1, \ldots, d\}^k$ be a multi-index. Let $h : \Delta_{t,s} \rightarrow \mathbb{R}$ be a bounded function. There exists $C > 0$ such that

\[
\sup_{x \in \mathbb{R}^d} \left| E[g(\bar{X}_{s,t}^{t,x})] \int_{t < t_1 < \cdots < t_k < s} h(t_1, \ldots, t_k) dW_{t_1}^{\alpha_1} \cdots dW_{t_k}^{\alpha_k} \right| \\
\leq C \varepsilon \#(\alpha; \alpha \not= 0) \left\| \nabla \#(\alpha; \alpha \not= 0) g \right\|_{\infty} (s - t)^k, \quad (14)
\]

for all $\varepsilon \in (0, 1]$, $g \in C^{\infty}_b(\mathbb{R}^d, \mathbb{R})$ and $t < s \leq T$.

**Proof of Lemma 1.** Use the duality formula in Malliavin calculus. $\square$

In the first step, we expand $P_{t,s}^\varepsilon \varphi$ for $\varphi \in C^{\infty}_b(\mathbb{R}^d)$ as follows:

\[
P_{t,s}^\varepsilon \varphi(x) = E[\varphi(\bar{X}_{s,t}^{t,x})] \\
+ \sum_{k=1}^{5} \sum_{\ell=1}^{k} \sum_{k_1+\cdots+k_{\ell}=k+\ell} \sum_{\alpha \in \{1, \ldots, d\}^\ell} \frac{1}{\ell!} \\
\left[ \partial_{\alpha(t)} \varphi(\bar{X}_{s,t}^{t,x}) \prod_{p=1}^{\ell} \varepsilon^{k_p} X_{t_p,\kappa_p,s}^{t,x,\alpha_p} \right] + \mathcal{A}^\varepsilon_{t,s} \varphi(x), \quad x \in \mathbb{R}^d.
\]

Here, $\mathcal{A}^\varepsilon_{t,s} \varphi$ satisfies $\sup_x \left| \mathcal{A}^\varepsilon_{t,s} \varphi(x) \right| \leq C \varepsilon^6 \left\| \varphi \right\|_{\infty} (s - t)^3$, where the constant $C > 0$ does not depend on $\varphi$ and $t, s$. Using
the integration by parts on the Wiener space with Lemma 1, we get
\begin{equation}
P_{t,T}^\varepsilon f(x) - Q_{t,t_1}^\varepsilon f(x) = \mathcal{M}_{1,\varepsilon}^t f(x) + \mathcal{M}_{1,\varepsilon}^t f(x), \quad x \in \mathbb{R}^d,
\end{equation}
where $\mathcal{M}_{1,\varepsilon}^t f(x)$ satisfies $\|\mathcal{M}_{1,\varepsilon}^t f(x)\| \leq C \sum_{\varepsilon > 0} \varepsilon^2 + \|\nabla f\| (s-t)^3$ for some $C > 0$ which does not depend on $\varepsilon$.

We now estimate the global error. Note that the following decomposition holds: for $x \in \mathbb{R}^d$,
\begin{equation}
P_{t,T}^\varepsilon f(x) - Q_{t,t_1}^\varepsilon f(x) \leq \sum_{i=0}^{n-1} Q_{t,t_1}^\varepsilon f(x) - Q_{t,t_{i+1}}^\varepsilon f(x) = \sum_{i=0}^{n-1} \left( Q_{t,t_1}^\varepsilon f(x) - Q_{t,t_{i+1}}^\varepsilon f(x) \right). \tag{16}
\end{equation}

For $Q_{t,t_1}^\varepsilon f(x) - Q_{t,t_{i+1}}^\varepsilon f(x)$, we have
\begin{equation}
Q_{t,t_1}^\varepsilon f(x) - Q_{t,t_{i+1}}^\varepsilon f(x) \leq C \varepsilon^2 \|f\| \|\mathcal{M}_{1,\varepsilon}^t f(x)\| \leq C \varepsilon^2 \|f\| \|\mathcal{M}_{1,\varepsilon}^t f(x)\| \leq C \varepsilon^2 \frac{1}{n^3}. \tag{17}
\end{equation}

We give the formula of Malliavin weight in the following:
\begin{equation}
W_s(x, \varepsilon) = 1 + \sum_{i=0}^{d} H_{i}(i, i, \varepsilon) A_{i}(i, t, x, s) \tag{18}
\end{equation}

APPENDIX

We give the formula of Malliavin weight in the following:
\begin{equation}
W_s(x, \varepsilon) = 1 + \sum_{i=0}^{d} H_{i}(i, i, \varepsilon) A_{i}(i, t, x, s) \tag{19}
\end{equation}

VI. CONCLUSIONS

In the paper, we introduced a computational scheme for diffusion semigroups based on an asymptotic expansion with weak approximation and deep learning algorithm to solve high-dimensional Kolmogorov PDEs. In particular, we provided a spatial approximation for the solution of $d$-dimensional PDEs on a hypercube $[a, b]^d$ without suffering from the curse of dimensionality. It can be regarded as an extension of classical finite element method of PDEs. Numerical experiments demonstrated the validity and the effectiveness of the proposed scheme.
ACKNOWLEDGEMENTS

We appreciate the comments from the two anonymous referees. The authors thank Prof. Shigeo Kusuoka (University of Tokyo) for grateful advice for the method of the paper. This work is supported by JST PRESTO (Grant Number JPMJPR2029), Japan.

REFERENCES


