Overcoming the course of dimensionality with Deep Learning: Methods and theoretical results for PDEs

Philippe von Wurstemberger (ETH Zurich, Switzerland)

Based on works by

Arnulf Jentzen (ETH Zurich, Switzerland),
Fabian Hornung (ETH Zurich, Switzerland & Karlsruhe Institute of Technology, Germany),
Philipp Grohs (University of Vienna, Austria),
Christian Beck (ETH Zurich, Switzerland),
Sebastian Becker (ZENAI AG, Switzerland),
Nor Jafaari (ZENAI AG, Switzerland),
Philippe von Wurstemberger (ETH Zurich, Switzerland)

ICCF 2019, A Coruna

11 July 2019

Goal: Approximate solutions of PDEs numerically, e.g., Black-Scholes PDE:

$$(\partial_t u)(t,x) + \sum_{i=1}^d \tfrac{|\beta_i|^2 |x_i|^2}{2} (\partial_{x_i x_i} u)(t,x) + \alpha_i x_i (\partial_{x_i} u)(t,x) = 0,$$

$$u(T,x) = g(x),$$

where $u\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$, $g\colon \mathbb{R}^d \to \mathbb{R}$, $(\alpha_i)_{i \in \{1,\ldots,d\}}, (\beta_i)_{i \in \{1,\ldots,d\}} \subseteq \mathbb{R}$, $T > 0$, $t \in [0,T]$, and $x \in \mathbb{R}^d$. Dimension $d \in \mathbb{N}$ corresponds to # assets in model.

Major issue for $d \gg 1$:

Classical approximations methods such as finite differences, finite element methods, sparse grids suffer under the **curse of dimensionality**.

Monte Carlo methods based on Feynman-Kac formula:
approximations at a fixed point without curse of dimensionality.

Methods based on Deep Learning :

- E, Han, & Jentzen, *Solving high-dimensional partial differential equations using deep learning. arXiv 2017. Proc. Natl. Acad. Sci. U.S.A. 2018*
- many more: E, Han, & Jentzen, Beck 2017; Becker, Grohs, Jaafari, Jentzen 2018; Fujii, Takahashi, Takahashi 2017; E, Yu 2017; Sirignano, Spiliopoulos 2017; Berg, Nyström 2017; Henry-Labordere 2017; Raissi 2018; Chan-Wai-Nam, Mikael, Warin 2018; Farahmand, Nabi, Nikovski 2017; Goudenege, Molent, Zanette 2019; Huré, Pham, Warin 2019; . . .

$\longrightarrow$ Is Deep Learning the key to **overcome** the **curse of dimensionality**?

Overview

A Deep Learning method for the Black-Scholes PDE

Error analysis

A Deep Learning method for the Black-Scholes PDE

Solving stochastic differential equations and Kolmogorov equations by means of deep learning, 2018 (Beck, Becker, Grohs, Jafaari, Jentzen)

1) Black-Scholes PDE as stochastic minimization problem

Let $d \in \mathbb{N}$, $(\alpha_i)_{i\in\{1,\dots,d\}}, (\beta_i)_{i\in\{1,\dots,d\}} \subseteq \mathbb{R}$, $T > 0$, $g \in C(\mathbb{R}^d, \mathbb{R})$, let $u \in C^{1,2}([0,T] \times \mathbb{R}^d, \mathbb{R})$ have at most poly. growing derivatives and satisfy $\forall\, t \in [0,T]$, $x \in \mathbb{R}^d$: $u(T,x) = g(x)$ and

$$(\partial_t u)(t,x) + \textstyle\sum_{i=1}^d \frac{|\beta_i|^2 |x_i|^2}{2} (\partial_{x_i x_i} u)(t,x) + \alpha_i x_i (\partial_{x_i} u)(t,x) = 0,$$

let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $W = (W^{(1)}, \dots, W^{(d)}) \colon [0,T] \times \Omega \to \mathbb{R}^d$ be a Brownian motion, for every $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ let $X(x) \colon \Omega \to \mathbb{R}^d$ satisfy

$$X(x) = \Big(x_1 e^{(\alpha_1 - \frac{\beta_1^2}{2})T + \beta_1 W_T^{(1)}}, \dots, x_d e^{(\alpha_d - \frac{\beta_d^2}{2})T + \beta_d W_T^{(d)}}\Big),$$

(Feynman-Kac Formula $\Rightarrow \forall x \in \mathbb{R}^d \colon u(0,x) = \mathbb{E}[g(X(x))]$)
let $a, b \in \mathbb{R}$, $a < b$, and let $\xi \colon \Omega \to [a,b]^d$ be $\mathcal{U}_{[a,b]^d}$-distributed indep. of W. Then

$$(u(0,\cdot))\big|_{[a,b]^d} = \operatorname*{argmin}_{v \in C([a,b]^d, \mathbb{R})} \mathbb{E}\Big[|v(\xi) - g(X(\xi))|^2\Big].$$

2) Rectified artificial neural network approximation of the minization problem

For all $n \in \mathbb{N}$, $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ let $\mathbf{A}_n \in C(\mathbb{R}^n, \mathbb{R}^n)$ satisfy $\mathbf{A}_n(x) = (\max\{x_1, 0\}, \ldots, \max\{x_n, 0\})$, let

$$\mathcal{N} = \cup_{L \in \{2,3,\ldots\}} \cup_{(l_0, l_1, \ldots, l_L) \in \mathbb{N}^{L+1}} \left(\times_{k=1}^{L} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}) \right),$$

let $\mathcal{R} \colon \mathcal{N} \to \cup_{k,l=1}^{\infty} C(\mathbb{R}^k, \mathbb{R}^l)$ satisfy $\forall\, L \geq 2$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$, $\phi = ((W_1, B_1), \ldots, (W_L, B_L)) \in (\times_{k=1}^{L} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}))$, $x \in \mathbb{R}^{l_0}$ that $\mathcal{R}(\phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}^{l_L})$ and

$$(\mathcal{R}(\phi))(x) = W_L(\mathbf{A}_{l_{L-1}}(\ldots \mathbf{A}_{l_2}(W_2(\mathbf{A}_{l_1}(W_1 x + B_1)) + B_2) \ldots)) + B_L.$$

Method

Fix ANN architecture $\mathcal{L} \geq 2$, $\mathfrak{l}_0 = d, \mathfrak{l}_1, \ldots, \mathfrak{l}_{\mathcal{L}-1}, \mathfrak{l}_{\mathcal{L}} = 1 \in \mathbb{N}$. Approximate

$$\phi^* = \operatorname*{argmin}_{\phi \in (\times_{k=1}^{\mathcal{L}} (\mathbb{R}^{\mathfrak{l}_k \times \mathfrak{l}_{k-1}} \times \mathbb{R}^{\mathfrak{l}_k}))} \mathbb{E}\left[|(\mathcal{R}(\phi))(\xi) - g(X(\xi))|^2 \right].$$

SGD approximations $(\Phi_n)_{n \in \mathbb{N}_0} \colon \mathbb{N}_0 \times \Omega \to \times_{k=1}^{\mathcal{L}} (\mathbb{R}^{\mathfrak{l}_k \times \mathfrak{l}_{k-1}} \times \mathbb{R}^{\mathfrak{l}_k})$.
For sufficiently large $n \in \mathbb{N}$:

$$(\mathcal{R}(\Phi_n))(\cdot) \approx \operatorname*{argmin}_{v \in C([a,b]^d, \mathbb{R})} \mathbb{E}\left[|v(\xi) - g(X(\xi))|^2 \right] = u(0, \cdot) \qquad \text{on} \quad [a,b]^d.$$

Error analysis

Goal:

- For suitable architecture and for large enough $n \in \mathbb{N}$:

$$\mathbb{E}\left[\|(\mathcal{R}(\Phi_n)) - u(0,\cdot)\|_{L^2([a,b]^d)}\right] < \varepsilon$$

- Investigate dependence on dimension

Approach:

- Divide the error

$$\begin{aligned}\mathbb{E}\big[\,&\|(\mathcal{R}(\Phi_n))(x) - u(0,x)\|_{L^2([a,b]^d)}\,\big]\\ &\le \underbrace{\mathbb{E}\big[\,\|(\mathcal{R}(\Phi_n)) - (\mathcal{R}(\phi^*))\|_{L^2([a,b]^d)}\,\big]}_{\text{Optimization error}} + \underbrace{\|(\mathcal{R}(\phi^*)) - u(0,\cdot)\|_{L^2([a,b]^d)}}_{\text{Approximation error}}\end{aligned}$$

- Optimization error: optimization of (highly) non-convex function

$\rightarrow$ No convergence guarantee (yet)

- Approximation error: This is what we want investigate

$\rightarrow$ What are ANNs capable of?

Theorem (Grohs, Hornung, Jentzen, vW 2018)

Let $T, \kappa > 0$, $(\alpha_{d,i})_{i\in\{1,\ldots,d\},d\in\mathbb{N}}, (\beta_{d,i})_{i\in\{1,\ldots,d\},d\in\mathbb{N}} \subseteq \mathbb{R}$ *satisfy that* $\sup_{i\in\{1,\ldots,d\},d\in\mathbb{N}}(|\alpha_{d,i}| + |\beta_{d,i}|) < \infty$, $\forall\, d \in \mathbb{N}$ *let* $g_d \in C(\mathbb{R}^d, \mathbb{R})$ *and let* $u_d \in C^{1,2}([0,T] \times \mathbb{R}^d, \mathbb{R})$ *be an at most poly. growing solution of*

$$(\partial_t u_d)(t,x) + \sum_{i=1}^d \tfrac{|\beta_{d,i}|^2 |x_i|^2}{2} (\partial_{x_i x_i} u_d)(t,x) + \alpha_{d,i} x_i (\partial_{x_i} u_d)(t,x) = 0,$$

$$u_d(T,x) = g_d(x),$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$, *let* $\mathcal{P} \colon \mathcal{N} \to \mathbb{N}$ *be the number of parameters, and let* $(\phi_{d,\delta})_{d\in\mathbb{N},\, \delta\in(0,1]} \subseteq \mathcal{N}$ *satisfy* $\forall\, d \in \mathbb{N}, \delta \in (0,1], x \in \mathbb{R}^d$ *that* $\mathcal{R}(\phi_{d,\delta}) \in C(\mathbb{R}^d, \mathbb{R})$, $|(\mathcal{R}(\phi_{d,\delta}))(x)| \le \kappa\, d^\kappa (1 + \|x\|_{\mathbb{R}^d}^\kappa)$, $\mathcal{P}(\phi_{d,\delta}) \le \kappa\, d^\kappa \delta^{-\kappa}$, *and* $|g_d(x) - (\mathcal{R}(\phi_{d,\delta}))(x)| \le \kappa\, d^\kappa\, \delta\, (1 + \|x\|_{\mathbb{R}^d}^\kappa)$.
Then $\exists\, c > 0$, $(\psi_{d,\varepsilon})_{d\in\mathbb{N},\, \varepsilon\in(0,1]} \subseteq \mathcal{N}$ *such that* $\forall\, d \in \mathbb{N}, \varepsilon \in (0,1]$: $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{P}(\psi_{d,\varepsilon}) \le c\, d^c\, \varepsilon^{-c}$, *and*

$$\left[\int_{[0,1]^d} |(\mathcal{R}(\psi_{d,\varepsilon}))(x) - u_d(0,x)|^2 \, dx\right]^{1/2} \le \varepsilon.$$

Example: $g_d(x_1, \ldots, x_d) = \max\left\{\frac{x_1 + \ldots + x_d}{d} - K, 0\right\}$ for some $K > 0$.

Main idea of the proof

- Statement purely *deterministic*. Proof based on *probabilistic arguments*.
- Use stochastic algorithm (Monte Carlo) which overcomes the course of dim. at single space points to construct ANN

→ Prove similar results for other PDEs

Jentzen, Salimova, Welti (2018):

$$(\partial_t u)(t,x) + \langle (\nabla_x u)(t,x), f(x) \rangle + \sum_{i,j=1}^{d} a_{ij} \, (\partial_{x_i x_j} u)(t,x) = 0, \qquad u(T,\cdot) = g$$

Hutzenthaler, Jentzen, Kruse, Nguyen (2019):

$$(\partial_t u)(t,x) + (\Delta_x u)(t,x) + f(u(t,x)) = 0, \qquad u(T,\cdot) = g$$

→ Theoretical (and practical) challenge: Find stochastic algorithms which overcome curse of dimensionality, e.g., Multilevel Picard Algorithm.

Recap

- Deep Learning methods for high-dim. PDEs seem to overcome curse of dim.
- A Deep Learning method for the Black-Scholes PDE
- Theoretical justification: ANNs have capacity to overcome curse of dim.
- Proof idea can be adapted for other PDEs

Thank you!

Appendix

- Proof of approximation capacity result for ANNs and PDEs
- Construction of random ANN for proof of approximation result
- Approximation result for nonlinear PDEs
- DNN Approximation result for nonlinear drift coefficient

Appendix: Proof of approximation capacity result for ANNs and PDEs

Main idea and comments about the proof

- Statement purely *deterministic*. Proof based on *probabilistic arguments*.
- Fix $d \in \mathbb{N}$, $\varepsilon \in (0,1]$. Goal: find $\psi \in \mathcal{N}$ such that $\mathcal{P}(\psi) \leq c\, d^c\, \varepsilon^{-c}$ and $\left[\int_{[0,1]^d} |u_d(0,x) - (\mathcal{R}(\psi))(x)|^2\, dx\right]^{1/2} \leq \varepsilon$.
- Main steps:
  - Construct random field $\mathcal{X} \colon \Omega \times \mathbb{R}^d \to \mathbb{R}$ with $\forall\, x \in \mathbb{R}^d \colon \mathcal{X}(x) \colon \Omega \to \mathbb{R} \sim$ Monte Carlo type approx of $u_d(0,x)$.
  - $\forall\, \omega \in \Omega \colon \mathcal{X}(\omega,\cdot) = \mathcal{R}(\Psi_\omega)$ is realization of ANN $\Psi_\omega \in \mathcal{N}$.
  - $\mathcal{X}$ is close to $u_d(0,\cdot)$: $\mathbb{E}\left[\left[\int_{[0,1]^d} |u_d(0,x) - \mathcal{X}(x)|^2\, dx\right]^{1/2}\right] \leq \varepsilon$.
  - $\Rightarrow \exists\, \boldsymbol{\omega} \in \Omega \colon \left[\int_{[0,1]^d} |u_d(0,x) - \underbrace{\mathcal{X}(\boldsymbol{\omega},x)}_{(\mathcal{R}(\Psi_{\boldsymbol{\omega}}))(x)}|^2\, dx\right]^{1/2} \leq \varepsilon$.
  - count parameters $\mathcal{P}(\Psi_{\boldsymbol{\omega}}) \leq c\, d^c\, \varepsilon^{-c}$.
  - define $\psi = \Psi_{\boldsymbol{\omega}}$.

Appendix: Construction of random ANN for proof of approximation result

Let $W^k\colon [0,T]\times\Omega\to\mathbb{R}^d$, $k\in\mathbb{N}$, be independent Brownian motions, let $X^k\colon\mathbb{R}^d\times\Omega\to\mathbb{R}^d$, $k\in\mathbb{N}$, be the random fields which satisfy $\forall\, x=(x_1,\ldots,x_d)\in\mathbb{R}^d\colon$

$$X^k(x)=\left[x_i\exp\!\left(\left(\alpha_{d,i}-\tfrac{|\beta_{d,i}|^2}{2}\right)T+\beta_{d,i}W_T^{d,\theta,i}\right)\right]_{i\in\{1,\ldots,d\}}.$$

Observe that $\forall\, x\in\mathbb{R}^d\colon\ u_d(0,x)=\mathbb{E}[g_d(X^1(x))]$.

We define $\mathcal{X}\colon\mathbb{R}^d\times\Omega\to\mathbb{R}$ by

$$\forall\, x\in\mathbb{R}^d\colon\qquad \mathcal{X}(x)=\frac{1}{M}\sum_{k=1}^{M}(\mathcal{R}(\phi_{d,\delta}))(X^k(x)),$$

for suitable $M\in\mathbb{N}$, $\delta\in(0,1]$.

Appendix: Approximation result for nonlinear PDEs

Theorem (Hutzenthaler, Jentzen, Kruse, Nguyen)

Let $T, \kappa > 0$, $f\colon \mathbb{R} \to \mathbb{R}$ *Lip. cont.,* $\forall\, d \in \mathbb{N}$ *let* $g_d \in C(\mathbb{R}^d, \mathbb{R})$ *and let* $u_d\colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$ *be an at most poly. grow. solution of*

$$\tfrac{\partial u_d}{\partial t} + \Delta_x u_d + f(u_d) = 0 \qquad \text{with} \qquad u_d(T, \cdot) = g_d,$$

and let $(\phi_{d,\varepsilon})_{d\in\mathbb{N},\, \varepsilon\in(0,1]} \subseteq \mathcal{N}$ *satisfy* $\forall\, d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$:
$\mathcal{R}(\phi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{P}(\phi_{d,\varepsilon}) \leq \kappa d^\kappa \varepsilon^{-\kappa}$, $|(\mathcal{R}(\phi_{d,\varepsilon}))(x)| \leq \kappa d^\kappa (1 + \|x\|^\kappa)$, *and*

$$|g_d(x) - (\mathcal{R}(\phi_{d,\varepsilon}))(x)| \leq \varepsilon \kappa d^\kappa (1 + \|x\|^\kappa_{\mathbb{R}^d}).$$

Then $\exists\, c \in (0, \infty)$, $(\psi_{d,\varepsilon})_{d\in\mathbb{N},\varepsilon\in(0,1]} \subseteq \mathcal{N}$ *such that* $\forall\, d \in \mathbb{N}$, $\varepsilon \in (0, 1]$:
$\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{P}(\psi_{d,\varepsilon}) \leq c\, d^c \varepsilon^{-c}$, *and*

$$\left[\int_{[0,1]^d} |u_d(0, x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^p \, dx\right]^{1/p} \leq \varepsilon.$$

Appendix: DNN Approximation result for nonlinear drift coefficient

Theorem (Jentzen, Salimova, & Welti 2018)

Let $T, \kappa, p > 0$, *let* $(a_{d,i,j})_{(i,j)\in\{1,\ldots,d\}^2} \in \mathbb{R}^{d\times d}$, $d \in \mathbb{N}$, *be sym. positive semi-def., let* $g_d\colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, *and* $f_d\colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, *be functions,* $\forall\, d \in \mathbb{N}$ *let* $u_d\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ *be at most poly. growing visc. solution of*

$$(\tfrac{\partial}{\partial t} u_d)(t,x) = (\tfrac{\partial}{\partial x} u_d)(t,x)\, f_d(x) + \sum_{i,j=1}^{d} a_{d,i,j}\, (\tfrac{\partial^2}{\partial x_i \partial x_j} u_d)(t,x), \, u_d(0,x) = g_d(x)$$

for $(t,x) \in (0,T)\times\mathbb{R}^d$, *and let* $(\phi^{m,d}_{\varepsilon})_{m\in\{0,1\},d\in\mathbb{N},\varepsilon\in(0,1]} \subseteq \mathcal{N}$ *satisfy* $\forall\, d \in \mathbb{N}, x, y \in \mathbb{R}^d\colon\; \mathcal{R}(\phi^{0,d}_{\varepsilon}) \in C(\mathbb{R}^d,\mathbb{R}), \mathcal{R}(\phi^{1,d}_{\varepsilon}) \in C(\mathbb{R}^d,\mathbb{R}^d)$,

$$|g_d(x)| + \textstyle\sum_{i,j=1}^{d} |a_{d,i,j}| \le \kappa d^{\kappa}(1 + \|x\|^{\kappa}_{\mathbb{R}^d}),$$

$$\|f_d(x) - f_d(y)\|_{\mathbb{R}^d} \le \kappa \|x-y\|_{\mathbb{R}^d}, \quad \|(\mathcal{R}\phi^{1,d}_{\varepsilon})(x)\|_{\mathbb{R}^d} \le \kappa(d^{\kappa} + \|x\|_{\mathbb{R}^d}),$$

$$|(\mathcal{R}\phi^{0,d}_{\varepsilon})(x) - (\mathcal{R}\phi^{0,d}_{\varepsilon})(y)| \le \kappa d^{\kappa}(1 + \|x\|^{\kappa}_{\mathbb{R}^d} + \|y\|^{\kappa}_{\mathbb{R}^d})\|x-y\|_{\mathbb{R}^d},$$

$$|g_d(x) - (\mathcal{R}\phi^{0,d}_{\varepsilon})(x)| + \|f_d(x) - (\mathcal{R}\phi^{1,d}_{\varepsilon})(x)\|_{\mathbb{R}^d} \le \varepsilon\kappa d^{\kappa}(1 + \|x\|^{\kappa}_{\mathbb{R}^d}),$$

and $\sum_{m=0}^{1} \mathcal{P}(\phi^{m,d}_{\varepsilon}) \le \kappa d^{\kappa}\varepsilon^{-\kappa}$. *Then* $\exists\, c \in (0,\infty), (\psi_{d,\varepsilon})_{d\in\mathbb{N},\varepsilon\in(0,1]} \subseteq \mathcal{N}$ *such that* $\forall\, d \in \mathbb{N}, \varepsilon \in (0,1]\colon\; \mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d,\mathbb{R}), \mathcal{P}(\psi_{d,\varepsilon}) \le c\, d^{c}\varepsilon^{-c}$, *and*

$$\left[\int_{[0,1]^d} |u_d(0,x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^p \, dy\right]^{1/p} \le \varepsilon.$$

Nonlinear equations?

Let $T > 0$, $\alpha, \beta \in \mathbb{R}$, $f \colon \mathbb{R} \to \mathbb{R}$ Lip. cont., $\forall\, d \in \mathbb{N}$ let $g_d \in C(\mathbb{R}^d, \mathbb{R})$ and let $u_d \in C([0,T] \times \mathbb{R}^d, \mathbb{R})$ an at most poly. growing visc. solution of

$$\frac{\partial u_d}{\partial t} + \left[\sum_{i=1}^{d} \frac{|\beta|^2 |x_i|^2}{2} \left(\frac{\partial^2 u_d}{\partial (x_i)^2}\right) + \alpha x_i \left(\frac{\partial u_d}{\partial x_i}\right)\right] + f(u_d) = 0,$$
$$u_d(T, x) = g_d(x),$$

for $(t, x) \in (0, T) \times \mathbb{R}^d$.

Application: e.g. pricing models with default risk (cf., e.g., Henry-Labordere 2012)

Full-history Recursive Multilevel Picard Algorithm (MLP)

(Hutzenthaler, Jentzen, Kruse, Nguyen, vW 2018)

Let $\kappa > 0$, $\Theta = \cup_{n=1}^{\infty} \mathbb{Z}^n$, $\forall\, d \in \mathbb{N}$, $x \in \mathbb{R}^d$ assume $|g_d(x)| \leq \kappa d^{\kappa}(1 + \|x\|_{\mathbb{R}^d}^{\kappa})$, let $(\Omega, \mathcal{F}, \mathbb{P})$ prob. space, let $W^{d,\theta} \colon [0,T] \times \Omega \to \mathbb{R}^d$, $\theta \in \Theta$, $d \in \mathbb{N}$, be i.i.d. Brownian motion, let $R^{\theta} \colon [0,T] \times \Omega \to [0,T]$, $\theta \in \Theta$, be i.i.d. cont. satisfying $\forall\, t \in [0,T], \theta \in \Theta$: R^{θ}_t is $\mathcal{U}_{[t,T]}$-distributed, assume $(W^{d,\theta})_{d \in \mathbb{N}, \theta \in \Theta}$ and $(R^{\theta})_{\theta \in \Theta}$ are indep., $\forall\, d \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ let $X^{d,\theta,x}_{t,s} \colon \Omega \to \mathbb{R}^d$ satisfy

$$X^{d,\theta,x}_{t,s} = \Big[x_i \exp\Big(\big(\alpha - \tfrac{\beta^2}{2}\big)(s-t) + \beta\big(W^{d,\theta,i}_s - W^{d,\theta,i}_t\big) \Big) \Big]_{i \in \{1,\ldots,d\}},$$

and let $V^{d,\theta}_{M,n} \colon [0,T] \times \mathbb{R}^d \times \Omega \to \mathbb{R}$, $M, n \in \mathbb{N}_0$, $\theta \in \Theta$, $d \in \mathbb{N}$, satisfy $\forall\, d, M, n \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that $V^{d,\theta}_{M,0}(t,x) = 0$ and

$$V^{d,\theta}_{M,n}(t,x) = \sum_{k=0}^{n-1} \tfrac{(T-t)}{M^{n-k}} \Bigg[\sum_{m=1}^{M^{n-k}} f\Big(V^{d,(\theta,k,m)}_{M,k}\Big(R^{(\theta,k,m)}_t, X^{d,(\theta,k,m),x}_{t,R^{(\theta,k,m)}_t}\Big)\Big) - \mathbb{1}_{\mathbb{N}}(k) f\Big(V^{d,(\theta,k,-m)}_{M,k-1}\Big(R^{(\theta,k,m)}_t, X^{d,(\theta,k,m),x}_{t,R^{(\theta,k,m)}_t}\Big)\Big) \Bigg] + \Bigg[\sum_{m=1}^{M^n} \frac{g_d(X^{d,(\theta,n,-m),x}_{t,T})}{M^n} \Bigg].$$

Then $\exists\, c > 0$, $N = (N_{d,\varepsilon})_{d\in\mathbb{N},\varepsilon\in(0,1]} \colon \mathbb{N} \times (0,1] \to \mathbb{N}$ such that $\forall\, d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $x \in [0,1]^d$:

$$\left(\mathbb{E}\left[\left|u_d(0,x) - V^{d,0}_{N_{d,\varepsilon},N_{d,\varepsilon}}(0,x)\right|^2\right]\right)^{1/2} \leq \varepsilon$$

and

$$\mathrm{Cost}\big(V^{d,0}_{N_{d,\varepsilon},N_{d,\varepsilon}}(0,x)\big) \leq c\, d^c \varepsilon^{-c}.$$

Corresponding approximation result for DNNs and semilinear PDEs:

- Hutzenthaler, Jentzen, Kruse, Nguyen, *A proof that rectified deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear heat equations. arXiv 2019.*

DNNs for Kolmogorov PDEs with nonlinear drift: Jentzen, Salimova, Welti 2018.