Overcoming the curse of dimensionality: from nonlinear Monte Carlo to the training of deep neural networks

Arnulf Jentzen (School of Data Science & Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong, Shenzhen, China & Institute of Analysis and Numerics, University of Münster, Germany)

Joint works with

Julia Ackermann (University of Wuppertal, Germany)
Sebastian Becker (ETH Zurich, Switzerland)
Steffen Dereich (University of Münster, Germany)
Thang Do (CUHK-Shenzhen, China &
Hanoi National University of Education, Vietnam)
Robin Graeber (University of Münster, Germany)
Lukas Gonon (Imperial College, UK)
Sonja Hannibal (University of Münster, Germany)
Martin Hutzenthaler (University of Duisburg-Essen, Germany)
Benno Kuckuck (University of Münster, Germany)
Thomas Kruse (University of Wuppertal, Germany)
Marvin Müller (2Xideas Switzerland AG, Switzerland)
Tuan Anh Nguyen (NTU Singapore, Singapore)
Joshua Lee Padgett (University of Arkansas, USA)
Adrian Riekert (University of Münster, Germany)
Philippe von Wurstemberger (CUHK-Shenzhen, China)

December 15, 2024

Funded by the European Union

European Research Council
Established by the European Commission

ANNs: For every $d \in \mathbb{N}$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ let

$$\mathcal{A}(x) = (\max\{x_1, 0\}, \ldots, \max\{x_d, 0\}),$$

let

$$\mathbf{N} = \cup_{L \in \mathbb{N}} \cup_{\ell_0, \ldots, \ell_L \in \mathbb{N}} \left(\times_{n=1}^{L} (\mathbb{R}^{\ell_n \times \ell_{n-1}} \times \mathbb{R}^{\ell_n}) \right),$$

and $\forall\, L \in \mathbb{N}$, $\ell_0, \ldots, \ell_L \in \mathbb{N}$, $U = ((W_1, B_1), \ldots, (W_L, B_L)) \in \times_{n=1}^{L} (\mathbb{R}^{\ell_n \times \ell_{n-1}} \times \mathbb{R}^{\ell_n})$ let $\mathcal{R}_U \colon \mathbb{R}^{\ell_0} \to \mathbb{R}^{\ell_L}$ and $\mathcal{P}_U \in \mathbb{N}$ satisfy for all $x_0 \in \mathbb{R}^{\ell_0}, \ldots, x_L \in \mathbb{R}^{\ell_L}$ with $\forall\, n \in \{1, \ldots, L\} \colon x_n = \mathcal{A}(W_n x_{n-1} + B_n)$ that

$$\mathcal{R}_U(x_0) = W_L x_{L-1} + B_L, \qquad \mathcal{P}_U = \textstyle\sum_{n=1}^{L} (\ell_n \ell_{n-1} + \ell_n).$$

Theorem (Hutzent.-J-Kruse-Nguyen 2020 *PDEA*; Ackerm.-Kruse-Kuck.-J-Padgett 2024)

Let $T, \kappa > 0$, *let* $f \colon \mathbb{R} \to \mathbb{R}$ *be Lipschitz,* $\forall\, d \in \mathbb{N}$ *let* $g_d \in C^1(\mathbb{R}^d, \mathbb{R})$ *and* $u_d \colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$ *be an at most poly. grow. solution of*

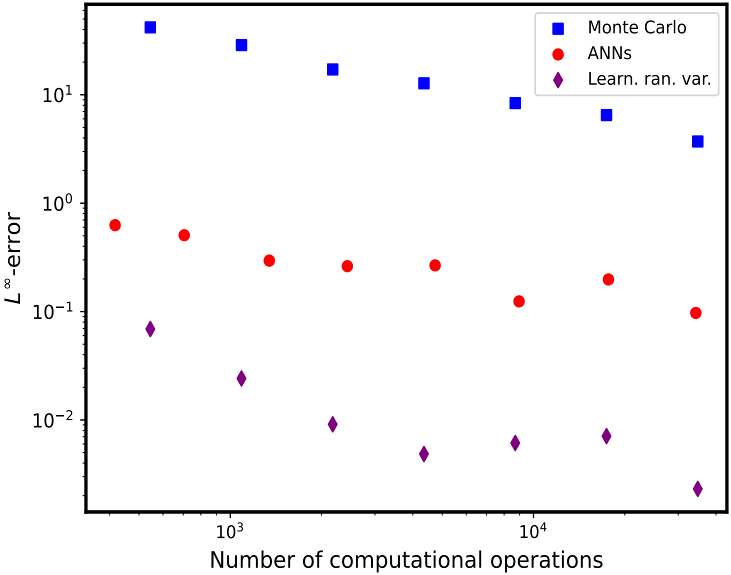
$$\tfrac{\partial u_d}{\partial t} = \Delta_x u_d + f(u_d), \qquad u_d(0, \cdot) = g_d$$

with $|g_d(x)| + \|(\nabla g_d)(x)\| \leq \kappa d^\kappa (1 + \|x\|^\kappa)$, *and assume* $\forall\, d \in \mathbb{N}, \varepsilon \in (0, 1] \colon \exists\, G \in \mathbf{N} \colon \forall\, x \in \mathbb{R}^d \colon |g_d(x) - \mathcal{R}_G(x)| \leq \varepsilon \kappa d^\kappa (1 + \|x\|^\kappa)$ *and* $\mathcal{P}_G \leq \kappa d^\kappa \varepsilon^{-\kappa}$.
Then $\exists\, c > 0 \colon \forall\, d \in \mathbb{N}, \varepsilon \in (0, 1] \colon \exists\, U \in \mathbf{N}$:

$$\int_{[0,T] \times [0,1]^d} |u_d(y) - \mathcal{R}_U(y)|\, dy \leq \varepsilon \qquad \text{and} \qquad \mathcal{P}_U \leq c\, d^c \varepsilon^{-c}.$$

Hutzenthaler-J-Kruse-Nguyen-von Wurstemberger 2020 *PRSA*,
Gonon-Graeber-J 2023, Becker-J-Müller-von Wurstemberger 2023 *Math. Finance*

$x \in [90, 110], \quad T \in [\frac{1}{100}, 1], \quad r \in [-\frac{1}{10}, \frac{1}{10}], \quad \sigma \in [\frac{1}{100}, \frac{1}{2}], \quad K \in [90, 110]$



(addition, substraction, multiplication, division, exponential, GeLU)

Theorem (Hannibal, J, & Thang 2024; J & Riekert 2024 SIAM JUQ (to appear))

Let $d \in \mathbb{N}$, $a \in \mathbb{R}$, $b > a$, *let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be prob. sp., for every* $m, n \in \mathbb{N}_0$ *let* $X_n^m \colon \Omega \to [a, b]^d$ *and* $Y_n^m \colon \Omega \to \mathbb{R}$ *be RVs, assume for all* $i \in \mathbb{N}$, $j \in \mathbb{N} \backslash \{i\}$ *that* $\mathbb{P}(X_0^i = X_0^j) = 0$, *for every* $k \in \mathbb{N}_0$ *let* $\mathfrak{d}_k, L_k \in \mathbb{N} \backslash \{1\}$, $\mathbf{l}_k = (\mathbf{l}_k^0, \ldots, \mathbf{l}_k^{L_k}) \in \mathbb{N}^{L_k+1}$ *satisfy* $\mathbf{l}_k^0 = d$, $\mathbf{l}_k^{L_k} = 1$, *and* $\mathfrak{d}_k = \sum_{i=1}^{L} \mathbf{l}_k^i (\mathbf{l}_k^{i-1} + 1)$, *for every* $k, n \in \mathbb{N}_0$, $\theta \in \mathbb{R}^{\mathfrak{d}_k}$ *let* $\mathcal{N}_k^{v,\theta} = (\mathcal{N}_{k,1}^{v,\theta}, \ldots, \mathcal{N}_{k,\mathbf{l}_k^v}^{v,\theta}) \colon \mathbb{R}^d \to \mathbb{R}^{\mathbf{l}_k^v}$, $v \in \{0, 1, \ldots, L_k\}$, *satisfy for all* $x \in \mathbb{R}^d$, $i \in \{1, \ldots, \mathbf{l}_k^{v+1}\}$ *that*

$$\mathcal{N}_{k,i}^{v+1,\theta}(x) = \theta_{\mathbf{l}_k^{v+1}\mathbf{l}_k^v + i + \sum_{h=1}^{v} \mathbf{l}_k^h(\mathbf{l}_k^{h-1}+1)} + \sum_{j=1}^{\mathbf{l}_k^v} \theta_{(i-1)\mathbf{l}_k^v + j + \sum_{h=1}^{v} \mathbf{l}_k^h(\mathbf{l}_k^{h-1}+1)} \big(x_j \mathbb{1}_{\{0\}}(v) + \max\{\mathcal{N}_{k,j}^{v,\theta}(x), 0\} \mathbb{1}_{\mathbb{N}}(v)\big),$$

for every $k, n \in \mathbb{N}_0$ *let* $M_n^k \in \mathbb{N}$, $\gamma_n^k \in \mathbb{R}$, *let* $\mathcal{R}_n^k \colon \mathbb{R}^{\mathfrak{d}_k} \times \Omega \to \mathbb{R}$ *satisfy*

$$\mathcal{R}_n^k(\theta) = \tfrac{1}{M_n^k} \Big[\sum_{m=1}^{M_n^k} |\mathcal{N}_k^{L_k,\theta}(X_n^m) - Y_n^m|^2\Big],$$

let $\mathfrak{G}_n^k \colon \mathbb{R}^{\mathfrak{d}_k} \times \Omega \to \mathbb{R}^{\mathfrak{d}_k}$ *be generalized gradient of* $\mathcal{R}_n^k$, *and let* $\Theta_n^k \colon \Omega \to \mathbb{R}^{\mathfrak{d}_k}$ *be RV, assume* $\forall\, k, n \in \mathbb{N} \colon \Theta_n^k = \Theta_{n-1}^k - \gamma_n^k \mathfrak{G}_n^k(\Theta_{n-1}^k)$, $\liminf_{k\to\infty} \mathbf{l}_k^1 = \infty$, *and* $\liminf_{k\to\infty} \mathbb{P}\big(\inf_{\theta \in \mathbb{R}^{\mathfrak{d}_k}} \mathcal{R}_0^k(\theta) > 0\big) = 1$, *let* $(c_k)_{k\in\mathbb{N}} \subseteq (0, \infty)$ *satisfy for all* $k \in \mathbb{N}$ *that* $c_k \Theta_0^k$ *is standard normal. Then*

$$\liminf_{k\to\infty} \mathbb{P}\Big(\inf_{n\in\mathbb{N}_0} \mathcal{R}_0^k(\Theta_n^k) > \inf_{\theta\in\mathbb{R}^{\mathfrak{d}_k}} \mathcal{R}_0^k(\theta)\Big) = 1.$$

Theorem (Dereich & J 2024: Convergence rates for Adam; Dereich, Graeber, & J 2024)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ *be prob. sp., let* $d \in \mathbb{N}$, $a \in \mathbb{R}$, $b > a$, *let* $X_{n,m} \colon \Omega \to [a, b]^d$, $(n, m) \in \mathbb{N}^2$, *be i.i.d. RVs, let* $\mathfrak{l} \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ *satisfy for all* $\theta, x \in \mathbb{R}^d$ *that*

$$\mathfrak{l}(\theta, x) = \|\theta - x\|^2,$$

let $\xi \in \mathbb{R}^d$, $p, \varepsilon, \alpha \in (0, \infty)$, $\beta \in (\alpha^2, 1)$, $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0, \infty)$ *satisfy* $\limsup_{n\to\infty}(\gamma_n + \gamma_n^{-2}|\gamma_n - \gamma_{n+1}|) = 0$, *for every* $M \in \mathbb{N}$ *let* $\mathcal{M}^M \colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$, $\mathbb{M}^M \colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$, *and* $\Theta^M \colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ *be stoch. proc. satisfying for all* $n \in \mathbb{N}$, $i \in \{1, \dots, d\}$ *that*

$$\mathcal{M}_0^M = 0, \qquad \mathcal{M}_n^M = \alpha \mathcal{M}_{n-1}^M + \frac{(1-\alpha)}{M}\left[\sum_{m=1}^M (\nabla_\theta \mathfrak{l})(\Theta_{n-1}^M, X_{n,m})\right],$$

$$\mathbb{M}_0^{M,i} = 0, \qquad \mathbb{M}_n^{M,i} = \beta \mathbb{M}_{n-1}^{M,i} + \frac{(1-\beta)}{M^2}\left[\sum_{m=1}^M \left(\tfrac{\partial}{\partial \theta_i}\mathfrak{l}\right)(\Theta_{n-1}^M, X_{n,m})\right]^2,$$

$$\Theta_0^M = \xi, \qquad \Theta_n^{M,i} = \Theta_{n-1}^{M,i} - \gamma_n\big[\varepsilon + [(1-\beta^n)^{-1}\mathbb{M}_n^{M,i}]^{1/2}\big]^{-1}\mathcal{M}_n^{M,i},$$

Then there exist $c > 0$, $(\vartheta_M)_{M\in\mathbb{N}} \subseteq \mathbb{R}^d$ *such that for all* $n \in \mathbb{N}$, $M \in \mathbb{N} \cap [c, \infty)$:

- $\mathbb{P}\big(\limsup_{N\to\infty} \|\Theta_N^M - \vartheta_M\| = 0\big) = 1$,
- $\big(\mathbb{E}[\|\Theta_n^M - \vartheta_M\|^p]\big)^{1/p} \le c\sqrt{\gamma_n}$, *and*
- $\|\vartheta_M - \mathbb{E}[X_{1,1}]\| \le c\, M^{-1}$.

Many thanks for your attention!

Funded by
the European Union

European Research Council
Established by the European Commission

Funded by the European Union. Views and opinions expressed are however those of the author only and do not necessarily reflect those of the European Union or European Research Council. Neither the European Union nor the granting authority can be held responsible for them.