Exponential Deep Neural Network Expression for Solution Sets of PDEs

Christoph Schwab

Seminar for Applied Mathematics

ETH Zürich, Switzerland

Joost Opschoor and Carlo Marcati (ETH),
Lukas Gonon (Munich), Lukas Herrmann (Linz), Philipp Petersen (Vienna), Jakob Zech (Heidelberg)

PSU WS on Neural Networks and PDEs, 16Dec, 2020

Motivation

- DNN **Approximation** and ML based **Computation** pervade all areas of CSE,
- **Hardware development** aligns with ML algorithms and data structures,
- **Data driven** simulation in **C**omputational (life, social, economic,...) **S**cience and **E**ngineering: "established" methodologies (FEM, FDM, FVM,...) suitable?
- **ML algorithms** offer **new paradigms** (adversarial NNs, self-play learning, ...) for numerical PDE solves,
- new role of "traditional" PDE solvers ("coaches" for DNN surrogates),
- UQ: **composition** of data ensemble model with PDE solution operator,
- new questions: given data, is PDE (model) right?

Implications

$\Longrightarrow$ **recast** established numerical DE simulation methodologies in the "DL world"
(PiNNs, ONets, quantized NNs and variable-precision numerics, ...)

$\Longrightarrow$ **analyze** mathematically the approximation capabilities of DL-based architectures for PDEs

$\Longrightarrow$ **synthesize strengths** of PDE-based and ML-based simulation methods.

$\Longrightarrow$ PDE simulation methods meet ML-dedicated hardware (TPUs, Neural Processors,...)

This talk: Deep ReLU NN Architectures emulate “best-in-class” numerical methods for

- AFEM
- hp-FEM
- Radial Bases
- Spectral Methods
- high-dimensional approximation (option pricing, UQ, Bayesian inverse problems,...)

Outline

1. Deep ReLU NN Approximation for PDE solution classes

- Deep ReLU and High Order FEM (mostly 1d) [Ph. Petersen, J. Opschoor]
  (a) ReLU DNN emulation of polynomials
  (b) ReLU DNN emulation of Splines
  (c) ReLU DNN emulation of hp-FEM (1d)
- Deep ReLU and hp-FEM (mostly 3d) [C. Marcati, Ph. Petersen, J. Opschoor]
  (a) weighted analytic regularity (countably normed spaces)
  (b) hp-FEM results
  (c) ReLU DNN expression rate bounds
- Deep ReLU expression rates of Option Prices in geometric Lévy Models (fractional parabolic) [L. Gonon]

2. Bayesian Inverse Problems (BIPs) [J. Opschoor, L. Herrmann, J. Zech]

- Examples: Diffusion Equation, Nonlinear Hyperbolic CL
- Prior constructions: Level Set, Affine-Parametric, Besov priors
- Holomorphy of Data-to-QoI map
- ReLU DNN Expression Rates

3. Conclusion, Extensions, References.

Deep ReLU Approximation

Activation of DNN: $\sigma(x) : \mathbb{R} \to \mathbb{R} : x \mapsto \max\{x, 0\}$.

- **Depth** $\sim$ *number of hidden layers* $L \in \mathbb{N}$,
- Numbers $N_\ell \in \mathbb{N}$ of *computation nodes in layer* $\ell \in \{0, \ldots, L\}$,

A *NN* Φ *with input dimension* d *and depth of* L *layers* is a sequence of matrix-vector tuples

$$\Phi = \big((W^1, b^1), (W^2, b^2), \ldots, (W^L, b^L)\big),$$

where $N_0 \coloneqq d$ and $N_1, \ldots, N_L \in \mathbb{N}$, and where $W^\ell \in \mathbb{R}^{N_\ell \times N_{\ell-1}}$ and $b^\ell \in \mathbb{R}^{N_\ell}$ for $\ell = 1, ..., L$.

Deep ReLU Approximation

ReLU DNN as a map $f : \mathbb{R}^{N_0} \to \mathbb{R}^{N_{L+1}}$ is **realized by DNN** Φ if ex. **weights** $w^\ell_{i,j} \in \mathbb{R}$, **biases** $b^\ell_j \in \mathbb{R}$ s.t. for all $x = (x_i)_{i=1}^{N_0}$ holds

$$z_j^1 = \sigma\left(\sum_{i=1}^{N_0} w_{i,j}^1 x_i + b_j^1\right) \;, \quad j \in \{1,\dots,N_1\} \;,$$

$$z_j^{\ell+1} = \sigma\left(\sum_{i=1}^{N_\ell} w_{i,j}^{\ell+1} z_i^\ell + b_j^{\ell+1}\right) \;, \quad \ell \in \{1,\dots,L-1\}, \;\; j \in \{1,\dots,N_{\ell+1}\} \;,$$

$$f(x) = (z_j^{L+1})_{j=1}^{N_{L+1}} = \left(\sum_{i=1}^{N_L} w_{i,j}^{L+1} z_i^L + b_j^{L+1}\right)_{j=1}^{N_{L+1}} \;.$$

N_0: dimension of NN input, N_{L+1}: dimension of NN output, $z_j^{\ell+1}$: output of unit j in layer ℓ.

$$\boxed{f = \mathrm{R}(\Phi)}$$

Deep ReLU Approximation

J. A. A. Opschoor and Ph. Petersen and ChS
Deep ReLU Networks and High-Order Finite Element Methods,
Anal. Appl. (Singap.) 18 (2020), no. 5, 715-770.

J. A. A. Opschoor and ChS and J. Zech
Exponential ReLU DNN expression of holomorphic maps in high dimension,
SAM Report 2019-35 (to appear in Constr. Approx.)

C. Marcati, J. A. A. Opschoor, Ph. Petersen and ChS
Exponential ReLU Neural Network Approximation Rates for Point and Edge Singularities
SAM Report 2020-65 (in review)
`https://math.ethz.ch/sam/research/reports.html?id=938`

Deep ReLU Approximation

Why ReLU?

- **Why not?**
- **Good numerical stability of ReLU DNN constructions,**
- **ReLU activation $\Longrightarrow \mathrm{R}(\Phi)$ is continuous,piecewise affine ("$\mathbb{P}_1$-FEM"),**
- **exact representation of rigid body motions,**
- **ReLU proofs are blueprints for corresponding arguments w. more sophisticated, *fixed* activations,**
  (!! (Substantial gains possible by "training the activation" [Maiorov&Pinkus]))
- **More Efficient Numerical Evaluation than e.g. $\tanh()$, sigmoid, softmax, etc.**

Deep ReLU Approximation

ReLU NNs and $\mathbb{P}_1$-FEM in $I = (0,1)$:

Lemma 1 ($\mathbb{P}_1$-FEM ReLU Emulation)**.**
For every partition $\mathcal{T}$ of $I = (0,1)$ with N elements and for every $v \in S_{\mathbf{1}}(I, \mathcal{T})$ ex. NN Φ^v such that

$$\mathrm{R}\left(\Phi^v\right) = v, \quad \mathrm{depth}\left(\Phi^v\right) = 2, \quad \mathrm{size}\left(\Phi^v\right) \leq 3N+1, \quad M_{\mathrm{fi}}\left(\Phi^v\right) \leq 2N, \quad M_{\mathrm{la}}\left(\Phi^v\right) \leq N+1. \tag{1}$$

Remark:

- Covers both, *fixed-knot* spline approximations, as well as *free-knot* spline approximations.

Corollary 2. *Assume*

- $s < \max\{2, 1+1/q\}$, $0 < q < q' \leq \infty$,
- $0 < s' < \min\{1+1/q', s-1/q+1/q'\}$, $0 < t, t' \leq \infty$.

Then ex. $C \coloneqq C(q, q', t, t', s, s') > 0$ s.t. for every $N \in \mathbb{N}$ and for every $f \in B^s_{q,t}(I)$ ex. NN Φ^N_f s.t.

$$\left\| f - \mathrm{R}\left(\Phi^N_f\right) \right\|_{B^{s'}_{q',t'}(I)} \leq C \left(\mathrm{size}\left(\Phi^N_f\right)\right)^{-(s-s')} \|f\|_{B^s_{q,t}(I)}.$$

Emulation of Polynomials by Deep ReLU NNs

Proposition 1 (Yarotsky 2017 ReLU Multiplication Lemma)**.**
Ex. $C_L, C'_L, C_M, C'_M > 0$ *such that, for every* $\kappa > 0$ *and* $\delta \in (0, 1/2)$*, ex. NN* $\widetilde{\times}_{\delta,\kappa}$ *s.t.*

$$\sup_{|a|,|b|\leq\kappa} \left| ab - \mathrm{R}\left(\widetilde{\times}_{\delta,\kappa}\right)(a,b) \right| \leq \delta \text{ and } \operatorname*{esssup}_{|a|,|b|\leq\kappa} \max\left\{ \left| a - \frac{\mathrm{d}}{\mathrm{d}b}\mathrm{R}\left(\widetilde{\times}_{\delta,\kappa}\right)(a,b) \right|, \left| b - \frac{\mathrm{d}}{\mathrm{d}a}\mathrm{R}\left(\widetilde{\times}_{\delta,\kappa}\right)(a,b) \right| \right\} \leq \delta,$$

$\mathrm{d}/\mathrm{d}a$ *and* $\mathrm{d}/\mathrm{d}b$ *weak derivatives.*

Furthermore, for every $\kappa > 0$ *and for every* $\delta \in (0, 1/2)$

$$M\left(\widetilde{\times}_{\delta,\kappa}\right) \leq C_M \left(\log_2\left(\frac{\max\{\kappa, 1\}}{\delta} \right) + 1 \right), \qquad L\left(\widetilde{\times}_{\delta,\kappa}\right) \leq C_L \left(\log_2\left(\frac{\max\{\kappa, 1\}}{\delta} \right) + 1 \right).$$

$$\forall a, b \in \mathbb{R}: \quad \mathrm{R}\left(\widetilde{\times}_{\delta,\kappa}\right)(a, 0) = \mathrm{R}\left(\widetilde{\times}_{\delta,\kappa}\right)(0, b) = 0. \tag{2}$$

Emulation of Polynomials by Deep ReLU NNs on $\hat{I} = (-1, 1)$

Proposition 2. *For each* $n \in \mathbb{N}_0$ *and each polynomial* $v \in \mathbb{P}_n([-1,1])$ *such that* $v(x) = \sum_{\ell=0}^{n} \tilde{v}_\ell x^\ell$*, with* $C_0 := \sum_{\ell=2}^{n} |\tilde{v}_\ell|$*, exist NNs* $\{\Phi_\varepsilon^v\}_{\varepsilon\in(0,1)}$ *with* $N_0 = N_l = 1$ *s.t.*

$$
\begin{aligned}
\|v - \mathrm{R}(\Phi_\varepsilon^v)\|_{W^{1,\infty}(\hat{I})} &\leq \varepsilon, \\
\mathrm{R}(\Phi_\varepsilon^v)(0) &= v(0), \\
\mathrm{depth}(\Phi_\varepsilon^v) &\lesssim (1 + \log_2(n)) \log_2(C_0/\varepsilon) + (\log_2(n))^3 \\
\mathrm{size}(\Phi_\varepsilon^v) &\lesssim n \log_2(C_0/\varepsilon) + n \log_2(n) + (1 + \log_2(n))^2 \log_2(C_0/\varepsilon)
\end{aligned}
$$

Emulation of hp-FEM by Deep ReLU NNs on $I = (0,1)$

Proposition 3. *For all $\boldsymbol{p} = (p_i)_{i\in\{1,\dots,N\}} \subset \mathbb{N}$, all partitions $\mathcal{T}$ of I into N open, disjoint, connected subintervals and for all $v \in S_{\boldsymbol{p}}(I,\mathcal{T})$, $0 < \varepsilon < 1/2$ ex. NNs $\{\Phi_\varepsilon^{v,\mathcal{T},\boldsymbol{p}}\}_{\varepsilon\in(0,1)}$ such that for all $1 \le q' \le \infty$ holds*

$$
\begin{aligned}
\left\| v - \mathrm{R}\left(\Phi_\varepsilon^{v,\mathcal{T},\boldsymbol{p}}\right)\right\|_{W^{1,q'}(I)} &\le \varepsilon \left|v\right|_{W^{1,q'}(I)}, \\
\mathrm{depth}\left(\Phi_\varepsilon^{v,\mathcal{T},\boldsymbol{p}}\right) &\lesssim \left(1+\log_2(p_{\max})\right)\left(2p_{\max}+\log_2(1/\varepsilon)\right) + \log_2(1/\varepsilon) + \left(1+\log_2^3(p_{\max})\right), \\
\mathrm{size}\left(\Phi_\varepsilon^{v,\mathcal{T},\boldsymbol{p}}\right) &\lesssim \sum_{i=1}^N p_i^2 + \log_2(1/\varepsilon)\sum_{i=1}^N p_i + \log_2(1/\varepsilon)\left(1+\sum_{i=1}^N \log_2^2(p_i)\right) \\
&\quad + \left(1+\sum_{i=1}^N p_i \log_2^2(p_i)\right) \\
&\quad + N\left(\left(1+\log_2(p_{\max})\right)\left(2p_{\max}+\log_2(1/\varepsilon)\right) + \left(1+\log_2^3(p_{\max})\right)\right).
\end{aligned}
$$

In addition, $\mathrm{R}\left(\Phi_\varepsilon^{v,\mathcal{T},\boldsymbol{p}}\right)(x_j) = v(x_j)$ for all $j \in \{0,\dots,N\}$, where $\{x_j\}_{j=0}^N$ are the nodes of $\mathcal{T}$.

Deep ReLU Approximation of Singularities

Emulation of hp-FEM by Deep ReLU NNs on $I = (0,1)$
Remark: ReLU DNNs Φ [whose realizations $\mathrm{R}(\Phi)$ are continuous, piecewise affine] deliver approximation rates of

- free-knot splines at fixed polynomial order, or free-knot, variable-degree splines,
- spectral methods and hp-FEM.

Singularities $u : I = (0,1) \to \mathbb{R}$ with point singularity at $x = 0$: **weighted analytic class**

$\mathcal{B}^{\ell}_{\beta}(I)$ is the set of all $u \in \bigcap_{k\geq\ell} H^{k,\ell}_{\beta}(I)$: exist $C_*(u), d(u) > 0$ such that

$$\boxed{\forall k \geq \ell : \quad |u|_{H^{k,\ell}_{\beta}(I)} \leq C_* d^{k-\ell}(k-\ell)! \ .}$$

Here

$$|u|_{H^{k,\ell}_{\beta}(I)} := \left\| x^{\beta+k-\ell} D^k u \right\|_{L^2(I)}, \quad \|u\|^2_{H^{k,\ell}_{\beta}(I)} := \begin{cases} \sum_{k'=0}^{k} |u|^2_{H^{k',0}_{\beta}(I)}, & \text{if } \ell = 0, \\ \sum_{k'=\ell}^{k} |u|^2_{H^{k',\ell}_{\beta}(I)} + \|u\|^2_{H^{\ell-1}(I)}, & \text{if } \ell \in \mathbb{N}. \end{cases}$$

Emulation of hp-FEM by Deep ReLU NNs on $I = (0,1)$

Theorem 4 (Exp. Convergence of hp-FEM)**.** *(K. Scherer, I. Babuška & B.Q.Guo, ...)*
Let $\sigma, \beta \in (0,1)$*,* $\lambda \coloneqq \sigma^{-1} - 1$*,* $u \in \mathcal{B}^2_\beta(I)$*.*

For $\mu_0 \coloneqq \mu_0(\sigma,\beta,d) \coloneqq \max\left\{1, \frac{d\lambda}{2\sigma^{1-\beta}}\right\}$ *and for* $\mu > \mu_0$ *let* $\boldsymbol{p} = (p_i)_{i=1}^N \subset \mathbb{N}$ *be defined as*

$$p_1 \coloneqq 1,\ p_i \coloneqq \lfloor \mu i \rfloor \text{ for } i \in \{2,\ldots,N\}.$$

Then ex. $C_7 > 0$ *and for every* $N \in \mathbb{N}$ *ex.* $v \in S_{\boldsymbol{p}}(I, \mathcal{T}_{\sigma,N})$ *such that*

- $v(x_i) = u(x_i)$ *for* $i \in \{1,\ldots,N\}$
- $$\|u - v\|_{H^1(I)} \le C_7 \exp\Big(-(1-\beta)\log(1/\sigma)N\Big) =: C_7 \exp(-cN).$$

As $N \to \infty$*,* $\dim(S_{\boldsymbol{p}}(I,\mathcal{T}_{\sigma,N})) = O(N^2)$ *i.e.*

$$\|u - v\|_{H^1(I)} \lesssim \exp(-bN_{DOF}^{1/2}).$$

Exponential Convergence of Deep ReLU NNs for singular functions on $I = (0, 1)$

Theorem 5. *[Opschoor, Petersen, ChS (AA 2020)]*
For all $\sigma, \beta \in (0,1)$, *all* $u \in \mathcal{B}^2_\beta(I)$ *and all* $\mu > \mu_0(\sigma, \beta, d(u))$
exist constants $C_8, c_9 > 0$ *and ReLU DNNs* $\{\Phi^{u,\sigma,N}\}_{N\in\mathbb{N}}$ *such that for all* $N \in \mathbb{N}$

$$\left\| u - \mathrm{R}(\Phi^{u,\sigma,N}) \right\|_{H^1(I)} \leq C_8 \exp(-c_9 N),$$

and such that

$$\mathrm{depth}(\Phi^{u,\sigma,N}) \lesssim N \log_2(N), \quad \mathrm{size}(\Phi^{u,\sigma,N}) \leq \left(2\mu^2 + \mu\right) N^3 + \left(1 + N^3 \log_2^2(N)\right),$$

Corollary: for every $\varepsilon > 0$ ex. $b = b(\varepsilon, \sigma, \beta, d(u))$ s.t. for all $N \in \mathbb{N}$

$$\left\| u - \mathrm{R}(\Phi^{u,\sigma,N}) \right\|_{H^1(I)} \lesssim \exp\left(-b\ (\mathrm{size}(\Phi^{u,\sigma,N}))^{1/3-\varepsilon}\right)$$

Analogous results for space dimension 2 and 3:

C. Marcati and J. A. A. Opschoor and P. C. Petersen and Ch. Schwab,

Exponential ReLU Neural Network Approximation Rates for Point and Edge Singularities

Report 2020-65, SAM, ETH `https://math.ethz.ch/sam/research/reports.html?id=938`

Radial Basis Functions (RBFs) in High Dimension [Opschoor, Petersen, ChS AA2020]

Proposition 6 ((B. McCane&L.Szymanski, Neurocomputing **313**(2018))). *For all dimensions $d \geq 2$ and for every target accuracy $\delta \in (0,1]$, exists a ReLU NN $\Phi^{\mathrm{Eucl}}_{d,\delta}$ with input dimension $N_0 = d$ and output dimension $N_L = 1$, such that $\mathrm{R}\left(\Phi^{\mathrm{Eucl}}_{d,\delta}\right)$ is 1-Lipschitz continuous,*

$$\left| \|x\|_{2,\mathbb{R}^d} - \mathrm{R}\left(\Phi^{\mathrm{Eucl}}_{d,\delta}\right)(x) \right| \leq \delta \|x\|_{2,\mathbb{R}^d}, \qquad \textit{for all } x \in \mathbb{R}^d, \tag{3}$$

$$\left| \|\cdot\|_{2,\mathbb{R}^d} - \mathrm{R}\left(\Phi^{\mathrm{Eucl}}_{d,\delta}\right) \right|_{W^{1,\infty}(\mathbb{R}^d)} \leq \delta, \qquad \textit{for a.e. } x \in \mathbb{R}^d, \tag{4}$$

and

$$\mathrm{depth}\left(\Phi^{\mathrm{Eucl}}_{d,\delta}\right) \leq \log_2(d)\log_2\left(10\pi\frac{d}{\delta}\right), \quad \mathrm{size}\left(\Phi^{\mathrm{Eucl}}_{d,\delta}\right) \leq 16(d-1)\log_2\left(10\pi\frac{d}{\delta}\right).$$

Radial Basis Functions (RBFs) in High Dimension [Opschoor, Petersen, ChS AA2020]

Theorem 7. *Let* $d \in \mathbb{N}$*,* $R > 0$*,* $A \in \mathbb{R}^{d\times d}$*,* $b \in \mathbb{R}^d$*, and* $D \coloneqq \{x \in \mathbb{R}^d : \|Ax + b\|_{2,\mathbb{R}^d} \le R\}$*.*

Let $g \in W^{2,\infty}([0, R])$ *be such that for all* $\beta \in (0, 1)$ *the function* g *can be approximated by a ReLU NN* $\Phi^g_{\beta,R}$ *with*

$$\left\| g - \mathrm{R}\left(\Phi^g_{\beta,R}\right) \right\|_{W^{1,\infty}([0,R])} \le \beta \, \|g\|_{W^{1,\infty}([0,R])} \,, \mathrm{depth}\left(\Phi^g_{\beta,R}\right) =: L_{\beta,R}, \quad \mathrm{size}\left(\Phi^g_{\beta,R}\right) =: M_{\beta,R}. \tag{5}$$

Consider **anisotropic radial-like function**

$$f : D \to \mathbb{R} : x \mapsto g(\|Ax + b\|_{2,\mathbb{R}^d}).$$

Then, for every $\varepsilon \in (0, 1]$ *exists a ReLU NN* $\Phi^f_{\varepsilon,D}$ *such that*

$$\left\| f - \mathrm{R}\left(\Phi^f_{\varepsilon,D}\right) \right\|_{L^\infty(D)} \le \varepsilon \, \|g\|_{W^{1,\infty}([0,R])} \,, \tag{6}$$

$$\left\| f - \mathrm{R}\left(\Phi^f_{\varepsilon,D}\right) \right\|_{W^{1,\infty}(D)} \le \varepsilon \, \|A\|_{2,\mathbb{R}^d} \, \|g\|_{W^{2,\infty}([0,R])} \,, \tag{7}$$

$$\mathrm{depth}\left(\Phi^f_{\varepsilon,D}\right) \le L_{\beta,R} + \log_2(d) \log_2\left(30\pi d\sqrt{d}\max\{R, 1\}/\varepsilon\right) + 1,$$

$$\mathrm{size}\left(\Phi^f_{\varepsilon,D}\right) \le 2M_{\beta,R} + 4d^2 + 64(d-1)\log_2\left(30\pi d\sqrt{d}\max\{R, 1\}/\varepsilon\right) + 4d.$$

Anisotropic RBF systems in dimension d can be emulated by ReLU DNNs
with exponential convergence **and *without curse of dimension.***

Option Pricing in Geometric Lévy Models in $\mathbb{R}^d$ [L. Gonon (Munich) and ChS (2020)]

$$u_d(\tau, s) = \mathbb{E}[\varphi_d(s_1 \exp(X^d_{\tau,1}), \ldots, s_d \exp(X^d_{\tau,d}))], \quad \tau \in [0,T], s \in (0,\infty)^d.$$

Assume that there exists constants $c > 0, p \geq 2, q \geq 0$ and, for all $\varepsilon \in (0,1]$, exists a ReLU NN $\varphi_{\varepsilon,d}$ with

$$|\varphi_d(s) - \mathrm{R}(\varphi_{\varepsilon,d})(s)| \leq \varepsilon c d^p (1 + \|s\|^p), \quad \text{for all } s \in (0,\infty)^d, \tag{8}$$

$$M(\varphi_{\varepsilon,d}) \leq c d^p \varepsilon^{-q}, \tag{9}$$

$$\mathrm{Lip}(\mathrm{R}(\varphi_{\varepsilon,d})) \leq c d^p. \tag{10}$$

In addition, assume that the Lévy triplets (A^d, γ^d, ν^d) of X^d are bounded:
exists $B \geq 0$ such that for each $d \in \mathbb{N}$, $i, j = 1, \ldots, d$,

$$A^d_{ij} \leq B,\ \gamma^d_i \leq B,\ \int_{\mathbb{R}^d \setminus \{\|y\| \leq 1\}} e^{p y_i} \nu^d(\mathrm{d}y) \leq B,\ \int_{\{\|y\| \leq 1\}} y_i^2 \nu^d(\mathrm{d}y) \leq B. \tag{11}$$

Then ex. constants $\kappa, \mathfrak{p}, \mathfrak{q} \in [0,\infty)$ and ReLU NNs $\psi_{\varepsilon,d}$, $\varepsilon \in (0,1]$, $d \in \mathbb{N}$ such that for any target accuracy $\varepsilon \in (0,1]$ and for any $d \in \mathbb{N}$ the number of weights grows only polynomially $M(\psi_{\varepsilon,d}) \leq \kappa d^{\mathfrak{p}} \varepsilon^{-\mathfrak{q}}$ and the approximation error between the NN $\psi_{\varepsilon,d}$ and the option price is at most ε, that is,

$$\sup_{s \in [a,b]^d} |u_d(T,s) - \mathrm{R}(\psi_{\varepsilon,d})(s)| \leq \varepsilon.$$

Deep ReLU network expression rates for option prices in high-dimensional, exponential Lévy models, L. Gonon and Ch. Schwab Report 2020-52, SAM

https://math.ethz.ch/sam/research/reports.html?id=925

Exponential ReLU DNN Expression of Electron Densities [MOPS 2020)]

$\Omega = \mathbb{R}^d/(2\mathbb{Z})^d$ periodic square/cube, $V : \Omega \to \mathbb{R}$ analytic potential s.t. $V(x) \geq V_0 > 0$ for all $x \in \Omega$

$$\exists \delta, A_V > 0 : \quad \|r^{2+|\alpha|-\delta}\partial^\alpha V\|_{L^\infty(\Omega)} \leq A_V^{|\alpha|+1}|\alpha|! \qquad \forall \alpha \in \mathbb{N}_0^d, \tag{12}$$

where $r(x) = \mathrm{dist}(x, (0, \ldots, 0))$.

Schrödinger eigenproblem: find smallest eigenvalue $\lambda \in \mathbb{R}$ and eigenfunction $u \in H^1(\Omega)$ s.t.

$$(-\Delta + V + |u|^k)u = \lambda u \quad \text{in } \Omega, \quad \|u\|_{L^2(\Omega)} = 1. \tag{13}$$

Proposition 8. *Let* $(\lambda, u) \in \mathbb{R} \times H^1(\Omega)$ *be a solution of problem* (13) *with minimal* λ*, where* V *satisfies* (12). *Then, for every* $\varepsilon \in (0, 1/2]$ *exists ReLU NN* $\Phi_{\varepsilon,u}$ *such that*

$$\|u - \mathrm{R}(\Phi_{\varepsilon,u})\|_{H^1(Q)} \leq \varepsilon, \tag{14}$$

and

$$\mathrm{size}(\Phi_{\varepsilon,u}) = \mathcal{O}(|\log_2(\varepsilon)|^{2d+1}), \quad \mathrm{depth}(\Phi_{\varepsilon,u}) = \mathcal{O}(|\log_2(\varepsilon)|\log_2(|\log_2(\varepsilon)|)).$$

[MOPS2020]: Exponential ReLU Neural Network Approximation Rates for Point and Edge Singularities

C. Marcati and J. A. A. Opschoor and P. C. Petersen and Ch. Schwab,

Report 2020-65, SAM, ETH `https://math.ethz.ch/sam/research/reports.html?id=938`

Bayesian Inverse Problems [M.Dashti & A.M.Stuart (2014)], [R.Nickl (2016)]

BIP: finite dimensional case
Goal:

Uncertain datum $u \in \mathbb{R}^n$ from **noisy observations** $\delta \in \mathbb{R}^K$.

Assume: $n \geq K$

- noiseless observable δ related to u by **forward map** $\delta = G(u)$, $u \mapsto G(u)$ Lipschitz
- δ accessible only up to **additive, centered observation noise**: centred RV $\eta \sim \mathbb{Q}_0$,
- Observation Noise Model:

$$\delta = G(u) + \eta \ , \quad \eta \sim \mathbb{Q}_0.$$

- Uncertain u: prior a.c. w.r. to Lebesgue measure

$$u \sim \pi_0 := \rho_0(u)\lambda^n \ .$$

- Observation Noise δ: law a.c. w.r. to Lebesgue measure λ^K in $\mathbb{R}^K$

$$\eta \sim \mathbb{Q}_0 = \rho\lambda^K \ .$$

Bayesian Inverse Problems[M.Dashti & A.M.Stuart (2014), R. Nickl (2016)]

BIP: finite dimensional case
Theorem [Bayes]
Assume **data** $\delta \in \mathbb{R}^K$ is such that

$$Z = Z(\delta) := \int_{\mathbb{R}^n} \rho(\delta - G(u))\rho_0(u)du > 0 \ .$$

Then, $u|\delta$ is RV on $\mathbb{R}^n$ distributed according to **Bayesian posterior** π^δ, with Lebesgue density

$$\rho^\delta(u) = \frac{1}{Z}\rho(\delta - G(u))\rho_0(u) \ , \quad u \in \mathbb{R}^n$$

w.r. to prior π_0 whose density with respect to λ^n on $\mathbb{R}^n$ is ρ_0.

Terminology:

- **likelihood**:

$$(u, \delta) \mapsto \rho(\delta - G(u))$$

- **Bayesian Potential**:

$$\Phi(u; \delta) := -\log(\rho(\delta - G(u)))$$

Bayesian Inverse Problems[M.Dashti & A.M.Stuart (2014), R. Nickl (2016)]

BIP: finite dimensional case
Example (Nondeg., centered gaussian obs. noise) $\mathbb{Q}_0 \sim \mathcal{N}(0, \Gamma)$, $\Gamma \in \mathbb{R}^{K\times K}_{>,sym}$

$$\Phi(u;\delta) = -\log \rho(\delta - G(u)) = \frac{1}{2}(\delta - G(u))^\top \Gamma^{-1}(\delta - G(u))$$

"negative log-likelihood", (observation noise) covariance-weighted "data-to-prediction misfit functional"

$$\frac{d\pi^\delta}{d\pi_0} = \frac{1}{Z}\exp(-\Phi(u;\delta)), \quad Z := \int_{u\in\mathbb{R}^n} \exp(-\Phi(u;\delta))\pi_0(du) \; .$$

Data-to-QoI ("Quantity of Interest") map: given, measurable QoI $\varphi : \mathbb{R}^n \to \mathbb{R}$, expectation under Bayesian posterior given data $\delta \in \mathbb{R}^K$, is

$$\mathbb{E}^{\pi^\delta}[\varphi] = \mathbb{E}^{\pi_0}\left[\frac{d\pi^\delta}{d\pi_0}\varphi\right] = \frac{1}{Z}\int_{u\in\mathbb{R}^n} \exp(-\Phi(u;\delta))\varphi(u)\pi_0(du) \; .$$

Goal:

express Data-to-QoI map $\delta \to \mathbb{E}^{\pi^\delta}[\varphi]$ **by DNN**.

Bayesian Inverse Problems[M.Dashti & A.M.Stuart (2014), R. Nickl (2016)]

X, Y real, sep. Banach, Borel σ-alg. $\sigma(X)$, $\sigma(Y)$.
$Y = \mathbb{R}^K$, X function space of uncertain PDE input.
Input-to-Observable map $G : X \to Y$ assumed measurable.
BIP: given obs. noise $\eta \in Y$, observe data δ such that

$$\delta = G(u) + \eta \ , \qquad u \in X \ .$$

Bayesian prior π_0 on $(X, \sigma(X))$, independent **noise model** $\eta \sim \mathbb{Q}_0$ on $(Y, \sigma(Y))$.

$\Longrightarrow u \sim \pi_0$, $\eta \sim \mathbb{Q}_0$ ind. RVs, $\nu_0 := \pi_0 \otimes \mathbb{Q}_0$ well defined.
Law $(u, \delta) \in X \times Y$: given $u \in X$, $\delta | u \in Y$ is RV with law $\mathbb{Q}_u$ (translate of $\mathbb{Q}_0$ by $G(u)$).
Assume

$$\mathbb{Q}_u << \mathbb{Q}_0 \quad \pi_0 - \text{a.e. } u \in X \ .$$

Then, for π_0-a.e. $u \in X$, $\Phi(u; \cdot) : Y \to \mathbb{R}$ measurable.
Law of RV (u, δ): $\quad (u, \delta) \sim \nu = \pi_0 \otimes \mathbb{Q}_u$, $\nu << \nu_0$ with

$$\frac{d\nu}{d\nu_0} = \exp(-\Phi(u; \delta)) \ .$$

Bayesian Inverse Problems[M.Dashti & A.M.Stuart (2014), R. Nickl (2016)]

Theorem [Bayes]

Assume $\Phi : X \times Y \to \mathbb{R}$ is ν_0-measurable with

$$\delta \mapsto Z(\delta) := \int_X \exp(-\Phi(u;\delta))\pi_0(du) > 0 \quad \mathbb{Q}_0\text{-a.e. } \delta \in Y \ .$$

Then law of $u|\delta$, π^δ, exists, and $\pi^\delta << \pi_0$. For ν-a.e. (u, δ)

$$\frac{d\pi^\delta}{d\pi_0} = \frac{1}{Z(\delta)} \exp(-\Phi(u;\delta)) \ .$$

PDE models

Setting:

- $u \in X' \subseteq X$ uncertain PDE input,
- $\mathcal{S} : X' \to V$ input-to-solution map, (Lipschitz) continuous,
- $\mathcal{O} \in (V^*)^K$ observation functional, $\quad G := \mathcal{O} \circ \mathcal{S} : X' \to Y = \mathbb{R}^K$ input-to-observable map,
- QoI $Q \in V^*$.
- Bayes: **Data-to-QoI map**

$$Y \ni \delta \mapsto \mathbb{E}^{\pi^\delta}[Q] = \frac{1}{Z(\delta)} \int_X (Q \circ \mathcal{S})(u)) \rho(\delta - G(u)) \pi_0(\mathrm{d}u) \in \mathbb{R}$$

  is **well-defined**.

- **Gaussian Obs. noise**: $\eta \sim \mathcal{N}(0, \Gamma) \implies$

$$Y \ni \delta \mapsto \mathbb{E}^{\pi^\delta}[Q] = \frac{1}{Z(\delta)} \int_X (Q \circ \mathcal{S})(u)) \exp(-\Phi(u; \delta)) \pi_0(\mathrm{d}u) \in \mathbb{R} \; .$$

PDE Example 1: Diffusion with uncertain coefficient

Diffusion: $D \subset \mathbb{R}^d$ bounded, Lipschitz, $f \in L^2(D)$ given.

$$f + \nabla \cdot (u \nabla q) = 0 \quad \text{in} \quad H^{-1}(D)\ ,\ q|_{\partial D} = 0\ .$$

- $V = H_0^1(D)$,
- input-to-solution map

$$\mathcal{S} : \{u \in L^\infty(D) : \operatorname{ess}\inf_{x \in D} u(x) > 0\} \to V : u \mapsto q$$

  satisfies

$$\|\mathcal{S}(u)\|_V \le \frac{\|f\|_{V^*}}{\operatorname{ess}\inf_{x \in D}\{u(x)\}}.$$

- $\mathcal{S}$ is Lipschitz continuous $\Longrightarrow$ measurability of the likelihood:
  $\forall u, u' \in \{u \in L^\infty(D) : \operatorname{ess}\inf_{x \in D} u(x) > 0\}$ such that $\mathcal{S}(u) \in W^{1,r}(D)$ for some $r \in [2, \infty)$,

$$\|\mathcal{S}(u) - \mathcal{S}(u')\|_V \le \frac{\|\nabla \mathcal{S}(u)\|_{L^r(D)}}{\operatorname{ess}\inf_{x \in D}\{u'(x)\}} \|u - u'\|_{L^{2r/(r-2)}(D)}\ .$$

  For $X' \subset \{u \in L^\infty : \operatorname{ess}\inf_{x \in D}\{u(x)\} > 0\}$ being (Borel) measurable, endow X' with the $L^\infty(D)$-norm and suppose that X' is separable with respect to the $L^\infty(D)$-norm, use $r' = 2r/(r-2) = \infty$ and $r = 2$.

PDE Example 2: SCL with uncertain flux

Nonlinear, scalar hyperbolic CL: Cauchy problem

$$\partial_t q + \partial_x(u(q)) = 0 \text{ in } \mathbb{R} \times \mathbb{R}_+, \qquad q|_{t=0} = q_0 \text{ in } \mathbb{R}\ .$$

Initial condition $q_0 \in L^1(\mathbb{R})$ has *bounded variation* and assumed known, i.e., deterministic.
Maximum principle satisfied by the (unique) entropy solutions implies:
Uncertain Lipschitz continuous flux function $u \in W^{1,\infty}([-M, M])$, $M := \|q_0\|_{L^\infty(\mathbb{R})}$.
Well-posed: For every flux $u \in W^{1,\infty}([-M, M])$ exists a unique *entropy solution* $q = \mathcal{S}_t(u)$ with

$$\forall t \geq 0: \qquad \|\mathcal{S}_t(u)\|_{L^1(\mathbb{R})} \leq \|q_0\|_{L^1(\mathbb{R})} + t\mathrm{TV}(q_0)\|\partial_x u\|_{L^\infty(-M,M)}.$$

Lipschitz-continuity of solution map:

$$\forall t \geq 0: \qquad \|\mathcal{S}_t(u) - \mathcal{S}_t(\tilde{u})\|_{L^1(\mathbb{R})} \leq t\mathrm{TV}(q_0)\|\partial_x(u - \tilde{u})\|_{L^\infty(-M,M)}.$$

$\Longrightarrow$ **Bayesian setting** with $X = W^{1,\infty}(-M, M)$, $V = L^1(\mathbb{R})$, $Q \in V^* = L^\infty(\mathbb{R})$.
Prior Constructions: Level-Set priors, Affine-parametric priors, Log-Besov parametric priors,...

[L. Herrmann, M. Keller and ChS:
Quasi-Monte Carlo Bayesian estimation under Besov priors in elliptic inverse problems.
MathComp 2020]

Regularity of the Data-to-QoI map

1. Lipschitz Regularity: finite-dimensional setting $u \in \mathbb{R}^n$, $\delta \in \mathbb{R}^K$:
Proposition[Herrmann, ChS, Zech (2020)]:
Assume

- $\rho \in \mathrm{Lip}(\mathbb{R}^K)$ with respect to norm $\| \circ \|$ on $\mathbb{R}^K$,
- QoI $Q \in L^1(X', \pi_0)$.

Then, for every $r > 0$ exists $C(r) > 0$ such that for every $\delta, \delta' \in \mathbb{R}^K$ with $\|\delta\|, \|\delta'\| \leq r$,

$$\forall \delta, \delta' \in B_r(0) : \quad \left| \mathbb{E}^{\pi^\delta}[\varphi] - \mathbb{E}^{\pi^{\delta'}}[\varphi] \right| \leq C \, \|\delta - \delta'\| \ .$$

2. Holomorphy: infinite-dimensional setting, $u \in X' \subseteq X$, $\delta \in \mathbb{R}^K$:
Proposition[Herrmann, ChS, Zech (2020)]:
Assume $Q \circ \mathcal{S} \in L^1(X', \pi_0)$ and $\Phi(u; \delta) = (\delta - G(u))^\top \Gamma^{-1}(\delta - G(u))/2$ for s.p.d. matrix $\Gamma \in \mathbb{R}^{K \times K}$.
Then **Data-to-QoI map**

$$\delta \mapsto \int_{X'} (Q \circ \mathcal{S})(u) \exp(-\Phi(u; \delta)) \pi_0(\mathrm{d}u)$$

is holomorphic .

ReLU DNN expression of the Data-to-QoI map

Theorem [Herrmann, ChS, Zech (2020)]
1. Forward UQ: Let $K \in \mathbb{N}$ and assume $f : \mathbb{C}^K \to \mathbb{C}$ is holomorphic and $f : \mathbb{R}^K \to \mathbb{R}$.
Then,
for all $\kappa > 1$, $r > 0$ ex. constant $C_{\kappa,r} > 0$ such that for all $n \in \mathbb{N}$
exists ReLU NN $\tilde{f}_n : [-r, r]^K \to \mathbb{R}$ such that

$$\sup_{\{\delta \in \mathbb{R}^K : |\delta| \leq r\}} |f(\delta) - \tilde{f}_n(\delta)| \leq C_{\kappa,r} \exp(-\kappa(\mathrm{size}(\tilde{f}_n))^{\frac{1}{K+1}}).$$

$$\mathrm{depth}(\tilde{f}_n) \leq C(1 + n\log(n)), \quad \mathrm{size}(\tilde{f}_n) \leq C(1+n)^{K+1} \text{ for } C > 0 \text{ ind. of } n.$$

2. Inverse UQ
Setting of inf. dimens. Bayes Theorem, centered **nondegenerate** gaussian obs. noise η.
Then, for all $r > 0$, $\kappa > 0$ exists $C_{\kappa,r} > 0$ such that $\forall n \in \mathbb{N}$ exists ReLU NN $\tilde{f}_n : [-r, r]^K \to \mathbb{R}$ s.t.

$$\sup_{\{\delta \in \mathbb{R}^K : |\delta| \leq r\}} \left| \mathbb{E}^{\pi^\delta}(\varphi) - \tilde{f}_n(\delta) \right| \leq C_{\kappa,r} \exp(-\kappa n).$$

Ex. $C_1 > 0$ s.t.

$$\mathrm{depth}(\tilde{f}_n) \leq C_1(1 + n\log(n)), \quad \mathrm{size}(\tilde{f}_n) \leq C_1(1+n)^{K+1}.$$

L. Herrmann, ChS, J. Zech: **Deep NN Expression Rates for posterior expectations in Bayesian Inversion** Inverse Problems (2020)

Conclusions

- Deep ReLU NNs emulate all major discretization schemes (h-FEM, Spectral FEM, hp-FEM, MsFEM, BEM, ENO, WENO, ...)
- Deep ReLU NNs allow **exponential convergence in terms of the DNN size** on solution sets of elliptic and parabolic PDEs with analytic data
- Deep NNs resolve in elliptic problems with multiple scales (homogenization, Helmholtz, ...) at **NN depth which is logarithmic in the scale parameter**
- DNNs emulate **variable float point precision algorithms through quantization and depth**
- ReLU DNNs **break the curse of dimensionality** in parametric PDEs in UQ
- DNNs represent solution manifolds for **high-dimensional, parametric PDEs** w.o. curse of dim.
- BIP: Sufficient conditions [additive gaussian obs. noise, Lipschitz forward solution map] for holomorphy of Data-to-QoI map in BIP; small ReLU NNs express Data-to-QoI very well.
- Exponential expression rates for Data-to-QoI map in BIP by DNN (and other architectures, sparse polynomials, tensors, ...)
- Proofs of emulation bounds ("in principle") constructive.

References

- **DNN expression rate analysis**

1. ChS and J. Zech:
   Deep learning in high dimension: Neural network expression rates for generalized polynomial chaos expansions in UQ,
   Analysis and Applications, Singapore, 17/1 (2019), pp. 19-55.
2. L. Herrmann, ChS, J. Zech:
   Deep ReLU NN Expression Rates for Data-to-QoI Maps in Bayesian Inversion
   Inverse Problems (2020)
3. J.A.A. Opschoor, Ph. Petersen and ChS:
   Deep ReLU Networks and High-Order FEM
   Analysis and Applications, Singapore (2020).
4. J. A. A. Opschoor and ChS and J. Zech:
   Exponential ReLU DNN expression of holomorphic maps in high dimension,
   SAM Report 2019-35 (to appear in Constr. Approx. 2021)

References

- **DNN expression rate analysis**

1. D. Elbrächter and Ph. Grohs and A. Jentzen and ChS:
   DNN Expression Rate Analysis of High-dimensional PDEs: Application to Option Pricing,
   (to appear in Constr. Approx. 2021).
2. V.H. Hoang and Ch. Schwab:
   Deep ReLU neural network expression for elliptic multiscale problems
   in review, SAM Report 2020-24,

References

- **Bayesian Inverse Problems**

1. M. Dashti and A. M. Stuart:
   The Bayesian approach to inverse problems.
   In: Handbook of uncertainty quantication. Vol. 3, pages 311-428.
   Springer, Cham, 2017.
2. V. H. Hoang, ChS, and A. Stuart:
   Complexity analysis of accelerated MCMC methods for Bayesian inversion.
   Inverse Problems, **29**(8), 2013.
3. V. H. Hoang, J. H. Quek, and ChS:
   Analysis of multilevel MCMC-FEM for Bayesian inversion of log-normal diffusions.
   Inverse Problems (2020)
   Report 2019-05, SAM, ETH Zürich.
4. L. Herrmann, M. Keller and ChS:
   Quasi-Monte Carlo Bayesian estimation under Besov priors in elliptic inverse problems.
   Math. Comp. (2020)

Thank You.