Some observations on high-dimensional partial differential equations with Barron data

Weinan E, Stephan Wojtowytsch

Princeton University

July 28, 2021

MSML 2021

Setting

PRINCETON UNIVERSITY

- There are many methods to solve 'low-dimensional' PDEs, but few for very high-dimensional problems (many body quantum systems in physics or computational chemistry, applications in finance, . . .)

Setting

- There are many methods to solve 'low-dimensional' PDEs, but few for very high-dimensional problems (many body quantum systems in physics or computational chemistry, applications in finance, ...)
- Neural networks have been very successful in beating the 'curse of dimensionality' in other fields and are quickly becoming more popular in solving high-dimensional PDEs as well
  - E-Han-Jentzen: *Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations*, 2017
  - E-Yu: *The Deep Ritz method: A deep learning-based numerical algorithm for solving variational problems*, 2017
  - Han-Jentzen-E: *Solving high-dimensional partial differential equations using deep learning*, 2017
  - ...

Setting

- There are many methods to solve 'low-dimensional' PDEs, but few for very high-dimensional problems (many body quantum systems in physics or computational chemistry, applications in finance, ...)
- Neural networks have been very successful in beating the 'curse of dimensionality' in other fields and are quickly becoming more popular in solving high-dimensional PDEs as well
  - E-Han-Jentzen: *Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations*, 2017
  - E-Yu: *The Deep Ritz method: A deep learning-based numerical algorithm for solving variational problems*, 2017
  - Han-Jentzen-E: *Solving high-dimensional partial differential equations using deep learning*, 2017
  - ...
- It can be proved rigorously in some cases that certain classes of neural networks can approximate the solution of a PDE without the CoD.

Setting

PRINCETON UNIVERSITY

- There are many methods to solve 'low-dimensional' PDEs, but few for very high-dimensional problems (many body quantum systems in physics or computational chemistry, applications in finance, ...)
- Neural networks have been very successful in beating the 'curse of dimensionality' in other fields and are quickly becoming more popular in solving high-dimensional PDEs as well
  - E-Han-Jentzen: *Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations*, 2017
  - E-Yu: *The Deep Ritz method: A deep learning-based numerical algorithm for solving variational problems*, 2017
  - Han-Jentzen-E: *Solving high-dimensional partial differential equations using deep learning*, 2017
  - ...
- It can be proved rigorously in some cases that certain classes of neural networks can approximate the solution of a PDE without the CoD.
- Can we show that solutions to PDEs lie in certain 'simple' function spaces, in which elements can be approximated by neural network models without the CoD?
  1. Regularity of solutions
  2. Approximation of elements in function space
  3. Numerical method

Setting

- There are many methods to solve 'low-dimensional' PDEs, but few for very high-dimensional problems (many body quantum systems in physics or computational chemistry, applications in finance, ...)
- Neural networks have been very successful in beating the 'curse of dimensionality' in other fields and are quickly becoming more popular in solving high-dimensional PDEs as well
  - E-Han-Jentzen: *Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations*, 2017
  - E-Yu: *The Deep Ritz method: A deep learning-based numerical algorithm for solving variational problems*, 2017
  - Han-Jentzen-E: *Solving high-dimensional partial differential equations using deep learning*, 2017
  - ...
- It can be proved rigorously in some cases that certain classes of neural networks can approximate the solution of a PDE without the CoD.
- Can we show that solutions to PDEs lie in certain 'simple' function spaces, in which elements can be approximated by neural network models without the CoD?
  1. **Regularity of solutions**
  2. Approximation of elements in function space
  3. Numerical method

PRINCETON UNIVERSITY

Definition

A function is called a *Barron function* if there exists a probability distribution π such that

$$f(x) = \mathbb{E}_{(a,w,b)\sim\pi}\left[a\sigma(w^T x + b)\right], \qquad \mathbb{E}_{(a,w,b)\sim\pi}\left[|a|(|w| + |b|)\right] < \infty.$$

Theorem (E-Ma-Wu '17)

If $\mathbb{P}$ is a probability distribution on $\mathbb{R}^d$ and $f : \mathbb{R}^d \to \mathbb{R}$ is a Barron function, then for every $m \in \mathbb{N}$ there exist weights $(a_i, w_i, b_i)_{i=1}^m$ such that

$$\left\| f - \sum_{i=1}^m a_i \sigma(w_i^T x + b_i) \right\|_{L^2(\mathbb{P})} \leq \frac{\|f\|_{\mathcal{B}}}{\sqrt{m}} \sqrt{\int_{\mathbb{R}^d} 1 + |x|^2 \, \mathrm{d}\mathbb{P}}.$$

Proof.

Monte-Carlo integration. $\square$

PRINCETON UNIVERSITY

Theorem

Assume that $f \in \mathcal{B}$ and

$$(-\Delta + \lambda^2) u = f$$

for $\lambda > 0$. Then $\|u\|_{\mathcal{B}} \leq \lambda^{-2} \, \|f\|_{\mathcal{B}}$ if σ has finite limits at $\pm\infty$ and

$$\|u\|_{\mathcal{B}} \leq \left[\lambda^{-2} + 2\,\lambda^{-3}\right] \|f\|_{\mathcal{B}}.$$

if $\sigma = \mathrm{ReLU}$.

Proof.

Convolution with the 3d Green's function $G(x) = \frac{e^{-\lambda|x|}}{4\pi\,|x|}$. Use that Barron functions are superpositions of 1d profiles. $\square$

Consider

$$\begin{cases} u_t - \Delta u &= 0 \quad & t > 0 \\ u &= u_0 \quad & t > 0 \end{cases}$$

With $u_0(x) = \mathbb{E}_{(a,w,b)}\left[a\sigma(w^T x + b)\right]$. Then

$$u(t,x) = \frac{1}{(4\pi)^{d/2}} \mathbb{E}_{(a,w,b)}\left[\int_{\mathbb{R}^d} \exp\left(-y^2\right) a\sigma\left(w^T\left(x - \sqrt{t}\, y\right) + b\right) \mathrm{d}y\right]$$

is a Barron function

1. in x and $\sqrt{t}$ with norm $\|u\|_{\mathcal{B}} \leq 2\,\|f\|_{\mathcal{B}}$.
2. in x with Barron norm $\|u(t,\cdot)\|_{\mathcal{B}} \leq 2(1+\sqrt{t})\,\|u\|_{\mathcal{B}}$ for fixed $t > 0$.

Remark

Only the second estimate carries over to the inhomogeneous heat equation as

$$\|u_{inhom}(t,\cdot)\|_{\mathcal{B}} \leq C\left[1 + t^{3/2}\right] \sup_{0<s<t} \|f(s,\cdot)\|_{\mathcal{B}}.$$

Viscous Hamilton-Jacobi equation

Consider

$$\begin{cases} u_t - \Delta u &= |\nabla u|^2 & t > 0 \\ u &= u_0 & t = 0 \end{cases}, \qquad \begin{cases} v_t - \Delta v &= 0 & t > 0 \\ v &= \exp(u_0) & t = 0. \end{cases}$$

Then $u(t,x) = \log\big(v(t,x)\big)$.

1. On bounded intervals, we can represent $\exp : (-\infty, R] \to \mathbb{R}$ and $\log : [\varepsilon, 1/\varepsilon] \to \mathbb{R}$ as Barron functions.

Viscous Hamilton-Jacobi equation

Consider

$$\begin{cases} u_t - \Delta u &= |\nabla u|^2 & t > 0 \\ u &= u_0 & t = 0 \end{cases}, \qquad \begin{cases} v_t - \Delta v &= 0 & t > 0 \\ v &= \exp(u_0) & t = 0. \end{cases}$$

Then $u(t,x) = \log\big(v(t,x)\big)$.

1. On bounded intervals, we can represent $\exp : (-\infty, R] \to \mathbb{R}$ and $\log : [\varepsilon, 1/\varepsilon] \to \mathbb{R}$ as Barron functions.
2. If $f = A \circ \sigma \circ W$ and $g = \tilde{A} \circ \sigma \circ \tilde{W}$ are finite two-layer networks, their composition

$$f \circ g = A \circ \sigma \circ \big(W \circ \tilde{A}\big) \circ \sigma \circ \tilde{W}$$

is a three-layer network, since the composition of linear maps is linear.

Viscous Hamilton-Jacobi equation

Consider

$$\begin{cases} u_t - \Delta u &= |\nabla u|^2 & t > 0 \\ u &= u_0 & t = 0 \end{cases}, \qquad \begin{cases} v_t - \Delta v &= 0 & t > 0 \\ v &= \exp(u_0) & t = 0. \end{cases}$$

Then $u(t,x) = \log\big(v(t,x)\big)$.

1. On bounded intervals, we can represent $\exp : (-\infty, R] \to \mathbb{R}$ and $\log : [\varepsilon, 1/\varepsilon] \to \mathbb{R}$ as Barron functions.
2. If $f = A \circ \sigma \circ W$ and $g = \tilde{A} \circ \sigma \circ \tilde{W}$ are finite two-layer networks, their composition

$$f \circ g = A \circ \sigma \circ \big(W \circ \tilde{A}\big) \circ \sigma \circ \tilde{W}$$

   is a three-layer network, since the composition of linear maps is linear.
3. Similarly, we can write $u = \phi \circ \exp \circ u_0$ where ϕ is a Barron function which includes the logarithm and the convolution with the heat kernel.

Viscous Hamilton-Jacobi equation

Consider

$$\begin{cases} u_t - \Delta u &= |\nabla u|^2 & t > 0 \\ u &= u_0 & t = 0 \end{cases}, \qquad \begin{cases} v_t - \Delta v &= 0 & t > 0 \\ v &= \exp(u_0) & t = 0. \end{cases}$$

Then $u(t,x) = \log\big(v(t,x)\big)$.

1. On bounded intervals, we can represent $\exp : (-\infty, R] \to \mathbb{R}$ and $\log : [\varepsilon, 1/\varepsilon] \to \mathbb{R}$ as Barron functions.
2. If $f = A \circ \sigma \circ W$ and $g = \tilde{A} \circ \sigma \circ \tilde{W}$ are finite two-layer networks, their composition
$$f \circ g = A \circ \sigma \circ \big(W \circ \tilde{A}\big) \circ \sigma \circ \tilde{W}$$
is a three-layer network, since the composition of linear maps is linear.
3. Similarly, we can write $u = \phi \circ \exp \circ u_0$ where ϕ is a Barron function which includes the logarithm and the convolution with the heat kernel.
4. In a function space $\mathcal{W}^3$ for three-layer neural networks (the space of 'tree-like' functions), the estimate
$$\|u\|_{\mathcal{W}^3} \leq \exp\left(\sup_{x \in \mathbb{R}^d} u_0(x) - \inf_{x \in \mathbb{R}^d} u_0(x)\right) \|u_0\|_{\mathcal{B}}$$
holds for bounded Barron initial data as a Barron function in $\sqrt{t}, x$.

Bounded domains

PRINCETON UNIVERSITY

Assume that u solves

$$\begin{cases} -\Delta u & = 0 & x \in B_1(0) \\ u & = \mathrm{ReLU}(x_1) & x \in \partial B_1(0). \end{cases}$$

Then

1. If $u \in \mathcal{B}$, it can be defined (non-uniquely) on the whole space by a representation formula $u(x) = \mathbb{E}_{(w,b)\sim\pi}\sigma(w^T x + b)$.

Bounded domains

Assume that u solves

$$\begin{cases} -\Delta u &= 0 & x \in B_1(0) \\ u &= \mathrm{ReLU}(x_1) & x \in \partial B_1(0). \end{cases}$$

Then

1. If $u \in \mathcal{B}$, it can be defined (non-uniquely) on the whole space by a representation formula $u(x) = \mathbb{E}_{(w,b)\sim\pi}\sigma(w^T x + b)$.
2. The equator $\{x \in \partial B_1(0) : x_1 = 0\}$ is part of the set where u is not differentiable, but the central plane $\{x \in B_1(0) : x_1 = 0\}$ is not.

Bounded domains

Assume that u solves

$$\begin{cases} -\Delta u &= 0 & x \in B_1(0) \\ u &= \mathrm{ReLU}(x_1) & x \in \partial B_1(0). \end{cases}$$

Then

1. If $u \in \mathcal{B}$, it can be defined (non-uniquely) on the whole space by a representation formula $u(x) = \mathbb{E}_{(w,b)\sim\pi}\sigma(w^T x + b)$.
2. The equator $\{x \in \partial B_1(0) : x_1 = 0\}$ is part of the set where u is not differentiable, but the central plane $\{x \in B_1(0) : x_1 = 0\}$ is not.
3. The singular set of a Barron function is a countable union of affine subspaces of dimension $0 \leq k \leq d - 1$ (structure theorem for Barron functions, E-W '20).

So, if $d > 2$, u is not a Barron function.

Observations

1. For translation-invariant linear PDEs, solutions often (but not always) lie in Barron space.
   - Examples: Heat equation, screened Poisson equation $-\Delta u + \lambda u = f$ for $\lambda > 0$.
   - Counterexample: Poisson equation, since the solution $u(x) = \frac{1}{6}\,\sigma(x_1)^3$ of $-\Delta u = \sigma(x_1)$ grows too fast at infinity.

Observations

1. For translation-invariant linear PDEs, solutions often (but not always) lie in Barron space.
    - Examples: Heat equation, screened Poisson equation $-\Delta u + \lambda u = f$ for $\lambda > 0$.
    - Counterexample: Poisson equation, since the solution $u(x) = \frac{1}{6}\sigma(x_1)^3$ of $-\Delta u = \sigma(x_1)$ grows too fast at infinity.
2. Boundary conditions break translation invariance. Even harmonic functions with ReLU boundary values are not in Barron space.
    - It is unknown whether the solution can be represented by e.g. the composition of Barron functions.
    - There are further positive results in this direction by Lu and Lu ('20, '21) for elliptic equations and Schrödinger eigenvalue problems with Neumann boundary condition and *spectral* Barron spaces on the unit hypercube.

Observations

1. For translation-invariant linear PDEs, solutions often (but not always) lie in Barron space.
    - Examples: Heat equation, screened Poisson equation $-\Delta u + \lambda u = f$ for $\lambda > 0$.
    - Counterexample: Poisson equation, since the solution $u(x) = \frac{1}{6}\sigma(x_1)^3$ of $-\Delta u = \sigma(x_1)$ grows too fast at infinity.
2. Boundary conditions break translation invariance. Even harmonic functions with ReLU boundary values are not in Barron space.
    - It is unknown whether the solution can be represented by e.g. the composition of Barron functions.
    - There are further positive results in this direction by Lu and Lu ('20, '21) for elliptic equations and Schrödinger eigenvalue problems with Neumann boundary condition and *spectral* Barron spaces on the unit hypercube.
3. Nonlinear PDEs may need deeper networks.