Numerical solution of Poisson-Neumann equation in the high dimensional regime using two-layer neural networks

Mathias Dus, Virginie Ehlarcher

Cermics - École des ponts ParisTech - Inria Paris team MATHERIALS

January 8, 2024

Summary

Poisson problem

Consider the Poisson-Neumann equation :

$$\begin{cases} -\Delta u^\star = f \text{ on } \Omega \\ \partial_n u^\star = 0 \text{ on } \partial\Omega \end{cases} \tag{1}$$

with $\Omega := [0,1]^d$ and f an L^2 source term.

Problem

How to solve numerically this problem when d is large ?

Why ? Quantum chemistry

The state is a wave function $\psi(x_1, \cdots, x_N)$ where $x_1, \cdots, x_N$ is the position of N particules ($x_i \in \mathbb{R}^3$). The fundamental state verifies :

$$\begin{cases} -\Delta\psi + V_n\psi + V_e\psi = E_0\psi \text{ on } \Omega \\ \partial_n\psi = 0 \text{ on } \partial\Omega \end{cases}$$

with :

- $V_n(x_1, \cdots, x_N) := -\sum_i^N \sum_j^M \frac{1}{|x_i - R_j|}$ (R_j position of atoms' nuclei)
- $V_e(x_1, \cdots, x_N) := \sum_i^N \sum_{j<i}^N \frac{1}{|x_i - x_j|}$

N can be very large *ie* 10, 100 $\rightarrow$ $d > 300$!

Why neural networks ?

- Some theory exists to reduce the dimension of the problem : DFT, Hartree-Fock, ... but it includes model approximations.
- Classical methods (finite volumes, finite elements,...) fail because of the use of a mesh $\rightarrow$ curse of dimensionality.

Some works show the relevance of using neural networks for regression problems and PDEs (FermiNet, PINN,...) in high dimension.

Barron functional space

We need an adapted space for neural networks [Lu et al., 2021, Barron, 1993].

- Spectral definition of $\mathcal{B}^s(\Omega)$ with $\mathcal{B} := \mathcal{B}^2$:

$$\|u\|_{\mathcal{B}^s} := \sum_{k\in\mathbb{N}^d} (1+\pi^s|k|_1^s)|\hat{u}(k)|$$

- The function u being written as :

$$u =: \sum_{k\in N^d} \hat{u}(k)\cos(\pi k_1 x_1)\cdots\cos(\pi k_d x_d)$$

Representation of a Barron function

Let $\chi : \mathbb{R} \to \mathbb{R}$ be an activation function and $B > 0$:

$$\mathcal{F}_\chi(B) := \left\{ \begin{array}{ll} x \to a\chi(w \cdot x + b) : & \\ a, b \in \mathbb{R}, & , |a| \leq 4B \\ w \in \mathbb{R}^d, & |w| = 1, |b| \leq 1 \end{array} \right\}$$

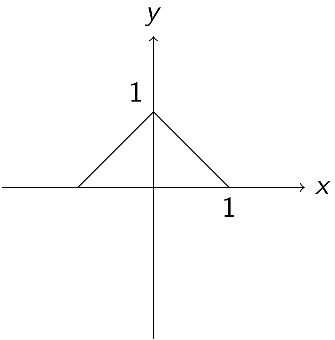
the space of features.

Lemma ([Lu et al., 2021])

If $u \in \mathcal{B}(\Omega)$ then u is the H^1 closure of the convex hull of $\mathcal{F}_{\cos}(\|u\|_{\mathcal{B}(\Omega)})$ and $\mathcal{F}_{\sigma_H}(\|u\|_{\mathcal{B}(\Omega)})$.

The hat activation function

$$\sigma_H(x) := \begin{cases} 0 & \text{si } |x| \geq 1 \\ x+1 & \text{si } -1 \leq x \leq 0 \\ 1-x & \text{si } 0 < x \leq 1. \end{cases}$$



Fundamental properties

Theorem ([Lu et al., 2021])

- *Let* $u \in \mathcal{B}(\Omega)$ *then for all* $m \in \mathbb{N}$*, there exists* $(a_i, w_i, b_i)_{i \leq m}$ *such that :*

$$\|u - u_m\|_{H^1(\Omega)} \leq C \frac{\|u\|_{\mathcal{B}(\Omega)}}{\sqrt{m}}$$

with $u_m(x) = \dfrac{1}{m} \displaystyle\sum_{i=1}^{m} a_i \sigma_H(w_i \cdot x + b_i)$.

- *If* $f \in \mathcal{B}^0(\Omega)$ *then the solution* $u^\star$ *of Poisson-Neumann's equation verifies* $\|u^\star\|_{\mathcal{B}(\Omega)} \leq d\|f\|_{\mathcal{B}^0(\Omega)}$

Numerical approximation results

Corollary

There exists $(a_i, w_i, b_i)_{i \leq m}$ *such that :*

$$\|u^\star - u_m\|_{H^1(\Omega)} \leq Cd \frac{\|f\|_{\mathcal{B}^0(\Omega)}}{\sqrt{m}}$$

with $u_m(x) = \int_\Theta a\sigma_H(w \cdot x + b) d\mu_m(a, w, b)$ *with :*

$$\mu_m := \frac{1}{m} \sum_{i=1}^{m} \delta_{a_i, w_i, b_i}$$

Neural network representation

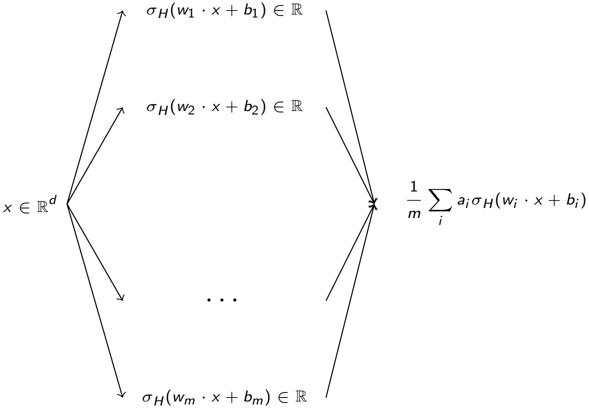


Figure: A two layer neural network representation

Fundamental idea and starting point

This suggests to study the following problem :

$$\mu^\star := \underset{\mu}{\operatorname{argmin}}\, \mathcal{E}_P(\mu) := \underset{\mu}{\operatorname{argmin}}\, \mathcal{E}\left(\int_\Theta a\sigma_H(w\cdot+b)d\mu(a,w,b)\right)$$

where $\mathcal{E}: H^1(\Omega) \to \mathbb{R}$ is the Poisson-Neumann energy :

$$\mathcal{E}(u) := \frac{1}{2}\left(\int_\Omega |\nabla u|^2 + \left(\int udx\right)^2\right) - \int_\Omega fudx.$$

We will consider the following gradient curve :

$$\boxed{\forall t \geq 0,\ \frac{d\mu_t}{dt} := -\nabla\,\mathcal{E}_P(\mu_t).}$$

Summary

Results

Theorem

There exists a gradient curve for the energy $\mathcal{E}_P$ and it satisfies the transport equation :

$$\begin{cases} \partial_t \mu_t + div((-\Pi v_{\mu_t})\mu_t) = 0 \\ \mu_{t=0} = \mu_0 \end{cases} \tag{2}$$

with $v_{\mu_t} := \nabla_\theta \phi_{\mu_t}$, $\phi_\mu(\theta) := d\,\mathcal{E}_{P_{\mathcal{B}(\mu)}}(\Phi(\theta;\cdot))$ and Π the projector onto the tangent space of Θ.

Results

Notation:

$$P_{\mathcal{B}}\mu := \int_{\Theta} a\sigma_H(w \cdot +b)d\mu(a,w,b)$$

the convex combination associated to the weight μ. The velocity potential writes :

$$\phi_\mu(\theta) := \int_\Omega \nabla(P_{\mathcal{B}}\mu)(x)\nabla\Phi(\theta;x) - f(x)\Phi(\theta;x)dx + \int_\Omega (P_{\mathcal{B}}\mu)(x)dx \int_\Omega \Phi(\theta;x)dx.$$

with $\theta := (a,w,b)$ and $\Phi(\theta;x) := a\sigma_H(w \cdot x + b)$.

Summary

Main theorem of convergence

Hypothesis (1)

The support of the initial measure μ_0 verifies :

$$\{0\} \times S_{\mathbb{R}^d}(1) \times [-\sqrt{d} - 2, \sqrt{d} + 2] \subset supp(\mu_0)$$

Theorem

Under hypothesis 1, if μ_t converges towards $\mu^\star$ in the Wasserstein sense then $\mu^\star$ is optimal.

Proof based on the work [Chizat and Bach, 2018] but there are differences...

Summary

Simulations

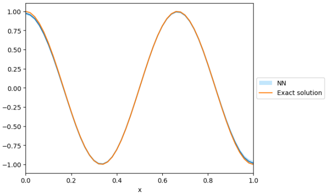
We consider modal solutions :

$$u_k(x_1, \cdots, x_d) := \cos(\pi k_1 \cdot x_1) \cdots \cos(\pi k_d \cdot x_d).$$

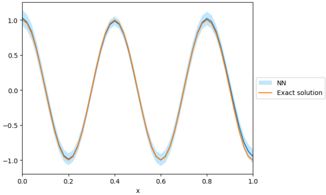
To know :

- Two-layer neural networks,
- integrals approximated by Monte-Carlo and automatic differentiation (tensorflow/keras with n samples per batch),
- gradient descent with learning rate equals to $\xi := \frac{1}{nm}$ with m the network's width.

In one dimension

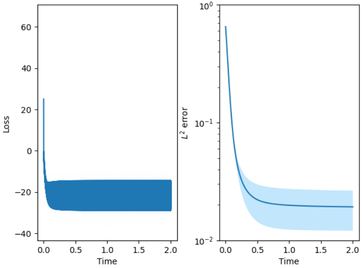


(a) $k = (3)$

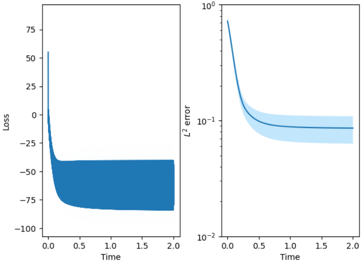


(b) $k = (5)$

Figure: Solution for $d = 1$ and $m = 1000$



(a) The case $d = 1$ and $k = (3)$

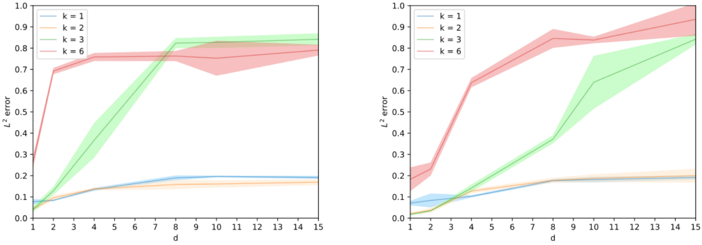


(b) The case $d = 1$ and $k = (5)$

Figure: Loss and L^2 error

Effects of dimension

$$d > 1, k = (k_1, 0, \cdots, 0),\ u_k(x_1, \cdots, x_d) := \cos(\pi k_1 \cdot x_1).$$



(a) $m = 100$ (b) $m = 1000$

Figure: Effects of dimension for different frequencies and width m

A case in high dimension

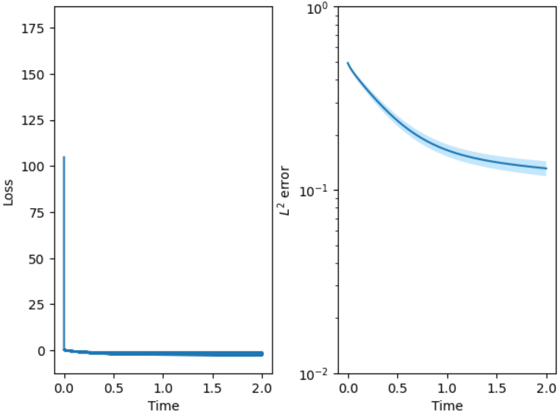


Figure: $m = 1000$, $k = (1, 1, 0, \cdots, 0)$, $d = 10$

A case in high dimension

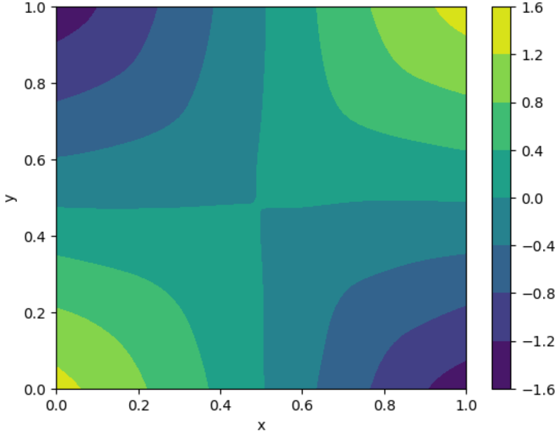


Figure: $m = 1000$, $k = (1, 1, 0, \cdots, 0)$, $d = 10$

What's next ?

- Quantum chemistry application,
- Dynamical PDEs $\rightarrow$ accelerate the learning phase,
- Multilayer neural network $\rightarrow$ promising results in simulation.

Bibliographie I

Ambrosio, L., Gigli, N., and Savaré, G. (2008).
Gradient flows in metric spaces and in the space of probability measures.
Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition.

Barron, A. R. (1993).
Universal approximation bounds for superpositions of a sigmoidal function.
IEEE Trans. Inform. Theory, 39(3):930–945.

Chizat, L. and Bach, F. (2018).
On the global convergence of gradient descent for over-parameterized models using optimal transport.

Bibliographie II

Lu, Y., Lu, J., and Wang, M. (2021).
A priori generalization analysis of the deep ritz method for solving high dimensional elliptic partial differential equations.
volume 134 of *Proceedings of Machine Learning Research*, pages 3196–3241. PMLR.

Ohta, S.-I. (2009).
Gradient flows on Wasserstein spaces over compact Alexandrov spaces.
Amer. J. Math., 131(2):475–516.