Numerical solution of elliptic equations in the high dimensional regime using two-layer neural networks

Mathias Dus, Virginie Ehrlacher

May 27, 2024

Cermics - École des ponts ParisTech - Inria Paris team MATHERIALS

Summary

Two problems

- 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.

- Consider the eigenvalue problem :

$$\begin{cases} -\Delta u^\star + W u^\star = E_0 u^\star \text{ on } \Omega \\ \partial_n u^\star = 0 \text{ on } \partial\Omega \end{cases} \tag{2}$$

where the potential W is "regular enough".

Problem

How to solve numerically these problems 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 $B > 0$:

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

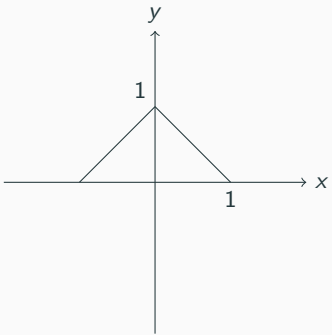
the space of features.

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

- *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 :*

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

The curse of dimensionality is broken for such regular functions !

For the eigenvalue problem, there is a result of the type : "If the potential W belongs to $\mathcal{B}$, then $u^\star$ belongs to $\mathcal{B}$" [Lu and Lu, 2022].

Neural network representation

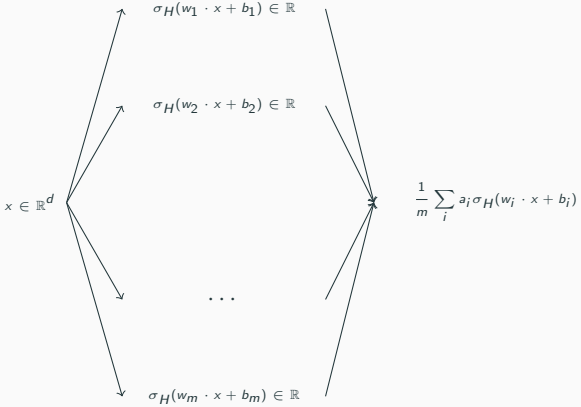


Figure 1: A two layer neural network representation

Fundamental idea and starting point

This suggests to study the following problem for the Poisson-Neumann case:

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

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

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

We will consider the following gradient curve :

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

Fundamental idea and starting point

The eigenvalue counterpart:

$$\mu^\star := \underset{\mu}{\operatorname{argmin}}\, \mathcal{E}_P(\mu)$$

contrained to the following condition:

$$C(\mu) := \left\| \int_\Theta a\sigma_H(w \cdot + b) d\mu(a, w, b) \right\|_{L^2(\Omega)} - 1 = 0.$$

We will consider the following stiff gradient curve :

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

with :

$$\mathcal{E}_P^C(\mu) = \begin{cases} \mathcal{E}_P(\mu) & \text{if } C(\mu) := 0 \\ +\infty & \text{otherwise .} \end{cases}$$

Summary

Results

Theorem

There exists a gradient curve for the energy and it satisfies the transport equation :

$$\begin{cases} \partial_t \mu_t + div((-\Pi v_{\mu_t})\mu_t) = 0 \\ \mu_{t=0} = \mu_0 \end{cases} \tag{3}$$

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.

Main ideas: Consider a regularized energy, prove its λ geodesic convexity wrt Wasserstein distance and use a bootstrap argument.

Results

- **Notation:**

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

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 (reminder) $\theta := (a, w, b)$ and $\Phi(\theta; x) := a\sigma_{H,\tau}(w \cdot x + b)$.

- For the eigenvalue problem, the velocity field writes $-\Pi(\nabla_\theta \phi_\mu - \sigma_\mu \nabla_\theta C_\mu)$ where: $C_\mu(\theta) := \dfrac{\langle P_\tau \mu, \Phi_\tau(\theta)\rangle_{L^2(\Omega)}}{\|P_\tau \mu\|^2_{L^2(\Omega)}}$ and $\sigma_\mu := \dfrac{\langle \nabla_\Theta \phi_\mu, \nabla_\Theta C_\mu\rangle_{L^2(\Theta;\mu)}}{\|\nabla_\Theta C_\mu\|^2_{L^2(\Theta;\mu)}}$ is a Lagrange multiplier.

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...

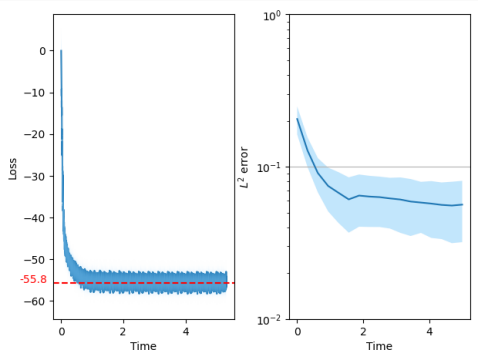
Summary

Simulations (for the eigenvalue case)

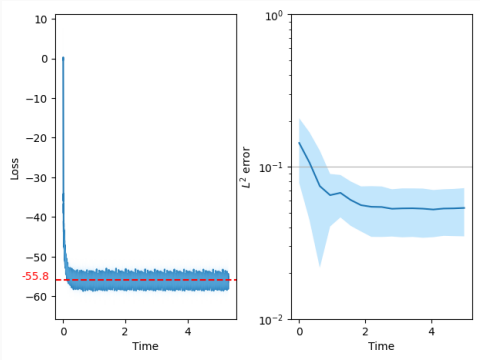
- We use the tensorflow/keras framework.
- Two-layer neural networks are used with a network width of either $m = 100$ or $m = 1000$.
- The dataset is made of 10^5 points sampled uniformly from the domain Ω.
- Batches are made of $n := 100$ or 1000 points taken from the dataset.
- The energy $\mathcal{E}$, the constraint C and its derivatives are computed by Monte-Carlo approximation and automatic differentiation.
- The optimizer is the classical stochastic gradient descent (SGD).
- At each time step, we normalize the last linear layer to impose the unit L^2 norm condition.
- To evaluate the performance of our method, we use a finite difference algorithm that is capable of computing eigenvectors and eigenvalues in the case where the potential depends only on the first two variables.
- Note that because of Monte-Carlo sampling, the neural network algorithm is stochastic. This is why, we ran our algorithms 8 times to evaluate the mean and the variance of our results.

Simulations: possible with tensors

$$W(x) = -100\cos(2\pi(x_1 - x_2))$$

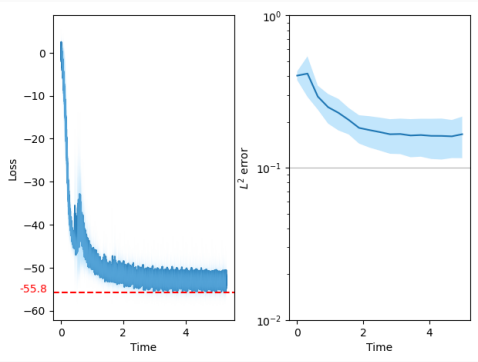


(a) $d = 2, m = 100$

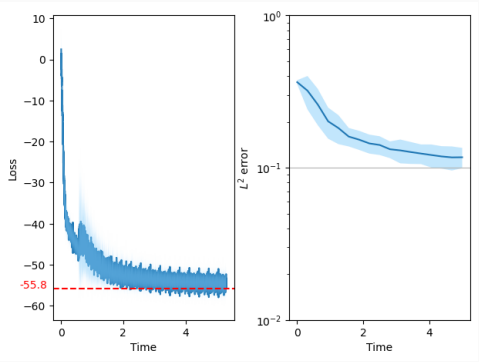


(b) $d = 2, m = 1000$

Simulations: possible with tensors



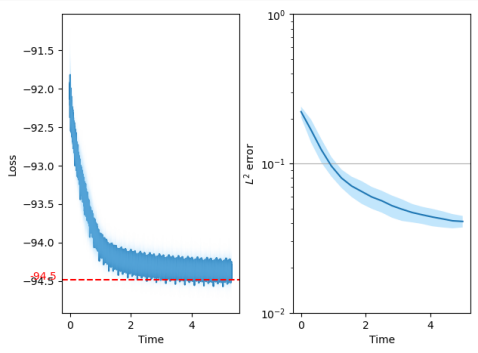
(a) $d = 8, m = 100$



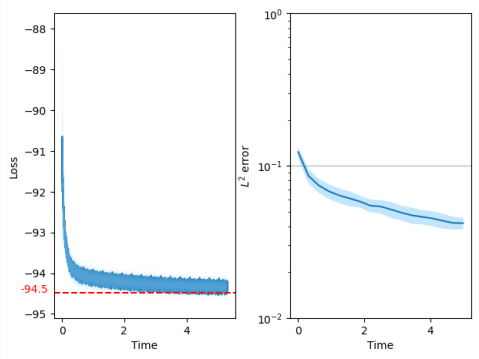
(b) $d = 8, m = 1000$

Simulations : not possible with tensors

$$W(x) = -100e^{-\frac{1}{2}(x_1 - x_2)^2}.$$



(a) $d = 8, m = 100$



(b) $d = 8, m = 1000$

Effects of n and m

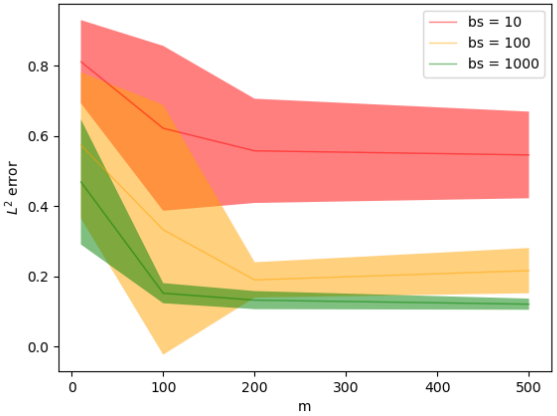


Figure 5: Effect of n and m

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.


Lu, J. and Lu, Y. (2022).
A priori generalization error analysis of two-layer neural networks for solving high dimensional Schrödinger eigenvalue problems.
Comm. Amer. Math. Soc., 2:1–21.

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.