Choosing the Right Ansatz Space: Towards Enhancing Random Feature Methods for Solving PDEs with Low Regularities and Constraints

Bowen Shi[1] and Yang Lyu[1]

[1]School of Mathematics, Peking University

Introduction

In recent years, machine learning methods, e.g. PINN, DRM, DGM has demonstrated success in solving partial differential equations and showed unique advantages over traditional methods. However, there are still problems concerning the accuracy of the solution in low dimensional cases:

1. Generally, optimization hardness makes accuracy not comparable to traditional methods. This is especially true when minimizing an adversial loss, which uses the global information of the PDE (weak form).
2. NN ansatz are chosen to be at least C^0 smooth, most times C^∞ smooth, but for most PDEs, solutions may not have such good regularities.
3. Some desirable properties of solution may not be preserved, e.g. conservation, symmetry, and loss of fine local structure. This will cause major accuracy problems.

The recent progress, random feature method for solving PDEs [3], can overcome problem 1, but problems 2 and 3 still needs to be better addressed. The key issue here is a misalignment of the function space of the solution and the function space of the ansatz.

- Novel viewpoint inspired by finite element analysis on improving accuracy of ML-based PDE solvers for PDE with low regularity solution or other special properties.
- Proposal of a *hp* adaptive method for Random Feature Method.
- An up-to-date account of the key relating literatures.

Illustrative Examples

Vector Laplacian on a Complex Domain

For a domain Ω in $\mathbb{R}^3$ with unit outward normal n, and $f:\mathbb{R}^3\to\mathbb{R}^3$, consider problem

$$-\operatorname{grad}\operatorname{div} u+\operatorname{curl}\operatorname{curl} u=f,\quad \text{in }\Omega,$$
$$u\cdot n=0,\quad \operatorname{curl} u\times n=0,\quad \text{on }\partial\Omega.$$

The solution u is the minimizer of the following energy functional:

$$J(u)=\frac{1}{2}\int_\Omega\left(|\operatorname{div} u|^2+|\operatorname{curl} u|^2\right)dx-\int_\Omega f\cdot u dx$$

Consider two situations.

(a) Ω is a nonconvex polyhedron. In this case, using theories of finite element analysis, we find solution is singular near the reentrant corner. However, any ansatz solution given by PINN, DGM, DRM or RFM will not have any regularity, which makes them differ greatly from true solution around that region.

(b) Ω is a multiple-connected region in $\mathbb{R}^3$. In this case, the solution is not unique. However, the solution space of PINN, DGM, DRM or RFM is a single-valued function space, which makes them unable to capture the true solution space.

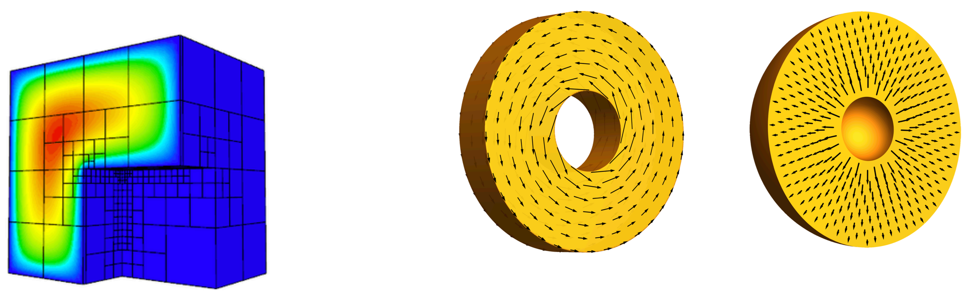


Figure 1: Illustrations for (a) and (b) respectively, taken from [1]

An accurate approximation of the vector Poisson equation should be obtained from a mixed formulation

$$\sigma=-\operatorname{div} u,\quad \operatorname{grad}\sigma+\operatorname{curl}\operatorname{curl} u=f \text{ in }\Omega,$$
$$u\cdot n=0,\quad \operatorname{curl} u\times n=0 \text{ on }\partial\Omega.$$

Here $\sigma\in H^1(\Omega), u\in H(\operatorname{curl};\Omega)$. $||u||$ needs to be minimized.

Related Works

Traditionally, **mixed finite element method** [1] can confine the ansatz space to satisfy various symmetries and regularities. However, it's a huge load of work to generate mesh.

Structure-preserving neural networks, e.g. **Hamiltonian neural network** [7], Lagrangian neural network [4], divergent free neural network [11], these methods use mathematical identity of the problem to smartly represent an ansatz in restricted space. However, they still face optimization problem and only applies to a restricted class of problems.

Discontinuous neural network [12, 5] is an approach for dealing with low regularity solution. By adding discontinuity at activation or weights, the network can represent discontinuous functions. However, the inclusion is not strict, and the discontinuity is not controlled. There are also work on learning adaptive collocation points for NN based solvers [10]

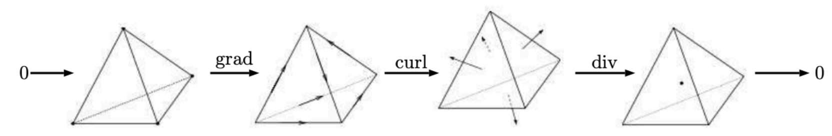
Theory of Solution Space

The function space of solution with various regularities and symmetries can be characterized by the inclusion relationships of certain Sobolev spaces. The ansatz space should be (at least an approximation of) a fininte dimensional subspace of the solution space. This is the key condition of accquring high accuracy. For the previous equation, we have:

Continous and Discrete de Rham Complex

$$\begin{array}{ccccccc}
H_0(\text{ grad }) & \xrightarrow{\text{grad}} & H_0(\text{ curl }) & \xrightarrow{\text{curl}} & H_0(\text{ div }) & \xrightarrow{\text{div}} & L_0^2\\
\downarrow\Pi_h^{\text{grad}} & & \downarrow\Pi_h^{\text{curl}} & & \downarrow\Pi_h^{\text{div}} & & \downarrow\Pi_h^0\\
H_0^h(\text{grad}) & \xrightarrow{\text{grad}} & H_0^h(\text{ curl }) & \xrightarrow{\text{curl}} & H_0^h(\text{ div }) & \xrightarrow{\text{div}} & L_0^{2,h}
\end{array}$$

Random Feature Function Complex



hp Adaptive Random Feature Method

Finer domain decomposition and more random feature functions are needed in regions with low regularity.

Mixed Random Feature Method

Consider the following problem after we have utilize the mixed formulation

$$\begin{cases}\mathcal{L}u(x)=f(x) & x\in\Omega\\ \mathcal{B}u(x)=g(x) & x\in\partial\Omega\end{cases}$$

where $x=(x_1,\cdots,x_d)^T$, and $\Omega=\Omega_1\cup\Omega_2\cdots\cup\Omega_m$

$$L(\Omega)=\sum_{\boldsymbol{x}_i\in C_I}\sum_{k=1}^{K_I}\lambda_{Ii}^k\left\|\mathcal{L}^k\boldsymbol{u}_M(\boldsymbol{x}_i)-\boldsymbol{f}^k(\boldsymbol{x}_i)\right\|_{l^2}^2$$
$$+\sum_{\boldsymbol{x}_j\in C_B}\sum_{\ell=1}^{K_B}\lambda_{Bj}^\ell\left\|\mathcal{B}^\ell\boldsymbol{u}_M(\boldsymbol{x}_j)-\boldsymbol{g}^\ell(\boldsymbol{x}_j)\right\|_{l^2}^2$$
$$+\text{ Conformity Loss}$$

C_I and C_B are interior points and boundary collocation points, respectively. $\boldsymbol{u}_M$ is the random feature basis, λ_{Ii}^k and λ_{Bj}^ℓ are hyperparameters. The optimization of the loss is done using standard convex optimization subroutine.

hp adaptivity

The main idea is the following loop:

$$\text{SOLVE}\rightarrow\text{ESTIMATE}\rightarrow\text{MARK}\rightarrow\text{REFINE}.$$

Starting from an initial PoU $\mathcal{T}_0$ to obtain PoU $\mathcal{T}_k$. Note that refinement method and posterior error has flexibility and may be chosen depends on the problem.

Algorithm 1 *hp* Adaptive Algorithm

```
Initialize T_0
for k = 0 to N do
    Solve to obtain u_k based on T_k (optimize loss in last section)
    Estimate e_k using u_k and T_k
    Mark a set of squares in T_k based on e_k
    Make T_{k+1} from T_k by partitioning marked squares into 4 small
    squares (h step)
    Add random feature basis to marked regions (p step)
end for
```

Error Analysis

An error analysis is important that it can give theoretical gurantee of the method and meanwhile gives a realization of the posterior error estimator. Random feature models are relatively simple compared to black box NNs and have been studied in different settings. [6, 8, 2] gave comprehensive generalization analysis of random feature methods versus two layer neural networks. [9] studied the error of random feature method in learning mappings between banach spaces, pointing out its intimate connection to kernel methods and Monte Carlo. In our problem the optimization error is negligible, so only generalization error and approximation error needs to be considered.

Conclusions

Choosing correct ansatz subspace is crucial for getting accurate solution. Adaptive method can make solution well approximate low regularity solution. Error analysis is both feasible and important for the problem.

Next Step

- An theoretical error analysis
- A numerical validation of proposed *hp* adaptive method
- Detailed examination of the effect of this subspace embedding
- Software package for solving real world problems

References

[1] Douglas N Arnold. *Finite element exterior calculus.* SIAM, 2018.

[2] Hongrui Chen, Jihao Long, and Lei Wu. A duality framework for generalization analysis of random feature models and two-layer neural networks. *arXiv preprint arXiv:2305.05642*, 2023.

[3] Jingrun Chen, Xurong Chi, Weinan E, and Zhouwang Yang. Bridging Traditional and Machine Learning-based Algorithms for Solving PDEs: The Random Feature Method, July 2022. arXiv:2207.13380 [physics].

[4] Miles Cranmer, Sam Greydanus, Stephan Hoyer, Peter Battaglia, David Spergel, and Shirley Ho. Lagrangian neural networks. *arXiv preprint arXiv:2003.04630*, 2020.

[5] Francesco Della Santa and Sandra Pieraccini. Discontinuous neural networks and discontinuity learning. *Journal of Computational and Applied Mathematics*, 419:114678, 2023.

[6] Weinan E, Chao Ma, and Lei Wu. A comparative analysis of optimization and generalization properties of two-layer neural network and random feature models under gradient descent dynamics. *Science China Mathematics*, 63(7):1235–1258, jan 2020.

[7] Samuel Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. *Advances in neural information processing systems*, 32, 2019.

[8] Chao Ma, Lei Wu, and Weinan E. The Slow Deterioration of the Generalization Error of the Random Feature Model. In *Proceedings of The First Mathematical and Scientific Machine Learning Conference*, pages 373–389. PMLR, August 2020. ISSN: 2640-3498.

[9] Nicholas H. Nelsen and Andrew M. Stuart. The Random Feature Model for Input-Output Maps between Banach Spaces. *SIAM Journal on Scientific Computing*, 43(5):A3212–A3243, January 2021. arXiv:2005.10224 [physics, stat].

[10] Ángel J. Omella and David Pardo. *r*Adaptive Deep Learning Method for Solving Partial Differential Equations, October 2022. arXiv:2210.10900 [cs, math].

[11] Jack Richter-Powell, Yaron Lipman, and Ricky T. Q. Chen. Neural Conservation Laws: A Divergence-Free Perspective, December 2022. arXiv:2210.01741 [cs].

[12] Jingbo Sun, Suchuan Dong, and Fei Wang. Local Randomized Neural Networks with Discontinuous Galerkin Methods for Partial Differential Equations, June 2022. arXiv:2206.05577 [cs, math].