Multilevel PINNs (MPINNs)

Valentin Mercier, Elisa Riccietti, Serge Gratton, Philippe Toint

IRIT / ENS Lyon / UNamur

May 20, 2022

Université Fédérale
Toulouse Midi-Pyrénées

The context

The problem: numerical approximation of PDE's solutions.

- *Classical approaches*: discretization and multigrid methods (MG)
- *New advances in machine learning*: Physics Informed Neural Networks (PINNs)

Our objective:

Transfer the advantages of the first approach to the second.

Outline

Multigrid methods

Physics Informed Neural Networks (PINNs)

Multilevel PINNs (MPINNs)

ANITI

Université Fédérale Toulouse Midi-Pyrénées

Outline

Multigrid methods

Physics Informed Neural Networks (PINNs)

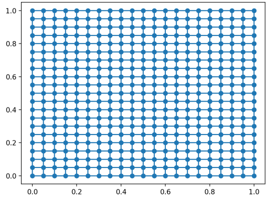
Multilevel PINNs (MPINNs)

ANITI

Université Fédérale Toulouse Midi-Pyrénées

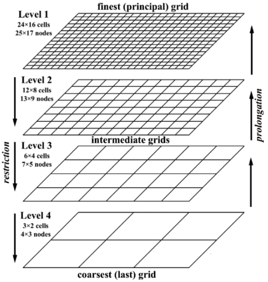
The numerical solution of PDEs

- Classically PDEs are discretized on a grid using finite differences or finite elements
- The resulting linear system $Au = f$ is solved using a fixed point iterative method (Gauss-Seidel or Jacobi)
- The size of the grids impacts the size of the system and the accuracy of the solution approximation



Multigrid methods for PDEs

State-of-the-art methods for PDEs: exploit representation of the problem at different scales



- Fine scales: eliminate high frequency components of the error
- Coarse scales: eliminate low frequency components of the error

ANITI

Université Fédérale Toulouse Midi-Pyrénées

The intuition behind multigrid methods

- A question of wavelength
- Example: consider $\Delta u = 0$, and take an initial guess consisting of the k-th Fourier mode $v_k(j) = \sin(\frac{kj\pi}{n})$

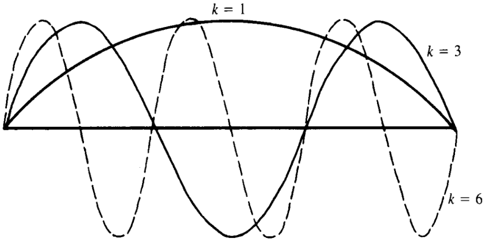


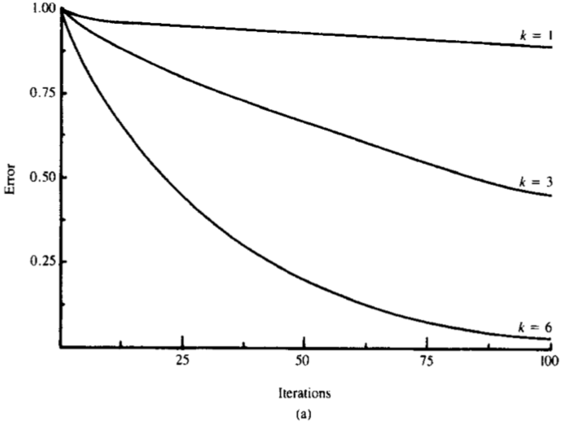
Figure 2.2: *The modes* $v_j = \sin\left(\frac{jk\pi}{n}\right)$, $0 \le j \le n$, *with wavenumbers* $k = 1, 3, 6$. *The kth mode consists of* $\frac{k}{2}$ *full sine waves on the interval.*

ANITI

Université Fédérale Toulouse Midi-Pyrénées

The intuition behind multigrid methods

- The *smoothing property*: hard for fixed point iterative methods to reduce the low frequency components of the error



The intuition behind multigrid methods

- How does a smooth component look like on a coarser grid?

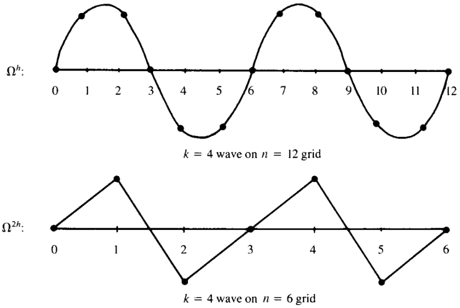


Figure 3.1: *Wave with wavenumber $k = 4$ on Ω^h ($n = 12$ points) projected onto Ω^{2h} ($n = 6$ points). The coarse grid "sees" a wave that is more oscillatory on the coarse grid than on the fine grid.*

ANITI

Université Fédérale Toulouse Midi-Pyrénées

Two-level multigrid methods

Consider a (possibly nonlinear) PDE:

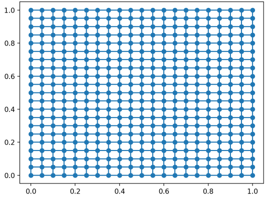
$$A(u) = f.$$

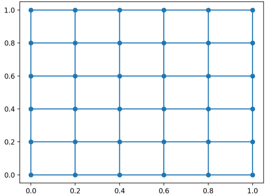
Consider two discretizations of the same system:

- Fine grid: $A_h(u_h) = f_h$
- Coarse grid: $A_H(u_H) = f_H$

Idea: write the solution u as the sum of a fine and a coarse term:

$$u \sim \underbrace{v_h}_{\in\mathbb{R}^h} + P(\underbrace{e_H}_{\in\mathbb{R}^H}), \; H < h.$$





Two-level multigrid methods

Build operators to transfer the information between the two levels

R:

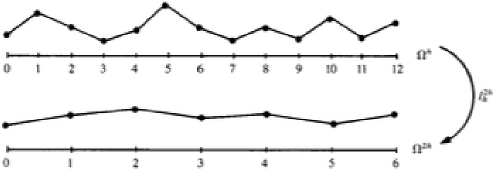
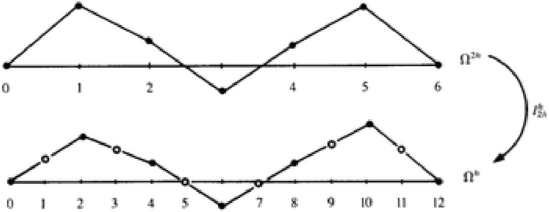


Figure 3.4: *Restriction by full weighting of a fine-grid vector to the coarse grid.*

P:



Two-level multigrid methods

Update the two components alternatively:

$$u \sim \underbrace{v_h}_{\in \mathbb{R}^h} + P(\underbrace{e_H}_{\in \mathbb{R}^H}), \; H < h.$$

- *Fine level*: get v_h by iterating on

$$A_h(u) = f_h$$

- *Coarse level*: compute correction by the residual equation:

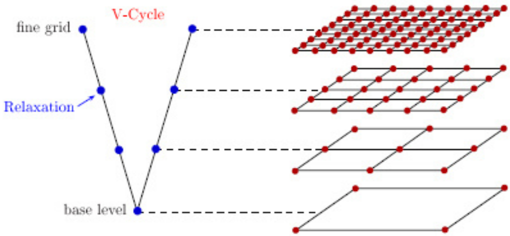
$$A_H(Rv_h + e_H) = A_H(Rv_h) + R(f_h - A_h(v_h))$$

- Correct: $v_h \leftarrow v_h + P(e_H)$

ANITI

Université Fédérale Toulouse Midi-Pyrénées

General multigrid methods



Optimal complexity for problems with diagonally dominant Fourier decomposition!

Outline

Multigrid methods

Physics Informed Neural Networks (PINNs)

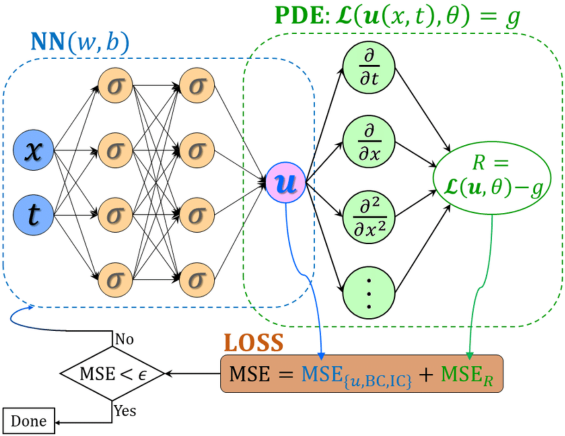
Multilevel PINNs (MPINNs)

ANITI

Université Fédérale Toulouse Midi-Pyrénées

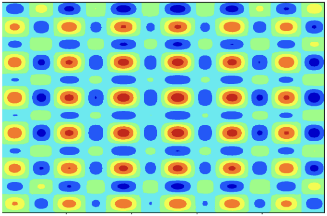
Physics Informed Neural Networks

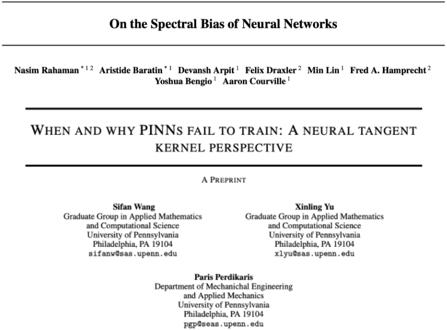
General principle



How does a PINN work?

On the spectral bias of neural networks





On the Spectral Bias of Neural Networks

Nasim Rahaman [*1,2] Aristide Baratin [*1] Devansh Arpit [1] Felix Draxler [2] Min Lin [1] Fred A. Hamprecht [2]
Yoshua Bengio [1] Aaron Courville [1]

WHEN AND WHY PINNS FAIL TO TRAIN: A NEURAL TANGENT KERNEL PERSPECTIVE

A PREPRINT

Sifan Wang
Graduate Group in Applied Mathematics
and Computational Science
University of Pennsylvania
Philadelphia, PA 19104
sifanw@sas.upenn.edu

Xinling Yu
Graduate Group in Applied Mathematics
and Computational Science
University of Pennsylvania
Philadelphia, PA 19104
xlyu@sas.upenn.edu

Paris Perdikaris
Department of Mechanichal Engineering
and Applied Mechanics
University of Pennsylvania
Philadelphia, PA 19104
pgp@seas.upenn.edu

$\Longrightarrow$ a standard single NN does not smooth the signal !

Université Fédérale
Toulouse Midi-Pyrénées

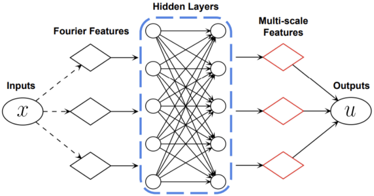
On the spectral bias of neural networks (1)

ON THE EIGENVECTOR BIAS OF FOURIER FEATURE NETWORKS: FROM REGRESSION TO SOLVING MULTI-SCALE PDES WITH PHYSICS-INFORMED NEURAL NETWORKS

Sifan Wang
Graduate Group in Applied Mathematics
and Computational Science
University of Pennsylvania
Philadelphia, PA 19104
sifanw@sas.upenn.edu

Hanwen Wang
Graduate Group in Applied Mathematics
and Computational Science
University of Pennsylvania
Philadelphia, PA 19104
wangh19@sas.upenn.edu

Paris Perdikaris
Department of Mechanichal Engineering
and Applied Mechanics
University of Pennsylvania
Philadelphia, PA 19104
pgp@seas.upenn.edu



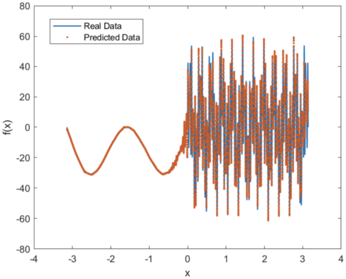
$$\boldsymbol{\gamma}^{(i)}(\boldsymbol{x}) = \begin{bmatrix} \cos(2\pi \boldsymbol{B}^{(i)}\boldsymbol{x}) \\ \sin(2\pi \boldsymbol{B}^{(i)}\boldsymbol{x}) \end{bmatrix}, \quad \text{for } i = 1, 2, \ldots, M$$

On the spectral bias of neural networks (2)

A PHASE SHIFT DEEP NEURAL NETWORK FOR HIGH FREQUENCY APPROXIMATION AND WAVE PROBLEMS *

WEI CAI†, XIAOGUANG LI‡, AND LIZUO LIU§

$$T(x) = \sum_{m=1}^{M} A_m \cos(\omega_m x) + B_m \sin(\omega_m x)$$



Université Fédérale Toulouse Midi-Pyrénées

Mscale networks (1)

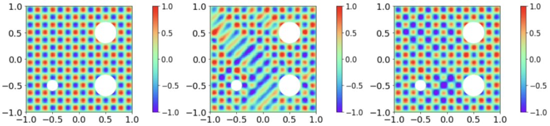
Multi-scale Deep Neural Network (MscaleDNN) for Solving Poisson-Boltzmann Equation in Complex Domains

Ziqi Liu[1], Wei Cai[2], and Zhi-Qin John Xu[3,*]

[1] *Beijing Computational Science Research Center, Beijing, 100193, PR China*
[2] *Dept. of Mathematics, Southern Methodist University, Dallas, TX 75275*
[3] *Institute of Natural Sciences, MOE-LSC and School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai, 200240, P.R. China*

$$f_i(\boldsymbol{x}) = \mathcal{F}^{-1}[\widehat{f}_i(\mathbf{k})](\boldsymbol{x}) = f(\boldsymbol{x}) * \chi^{\vee}_{A_i}(\boldsymbol{x}),$$

$$f(\boldsymbol{x}) \sim \sum_{i=1}^{M} f_{\theta^{n_i}}(\alpha_i \boldsymbol{x}).$$

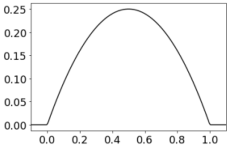


(a) exact (b) normal (c) Mscale

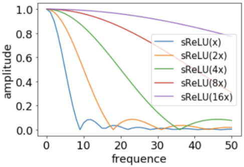
Figure 17: Exact and numerical solution for the Poisson equation in domain 1.

Mscale networks (2)

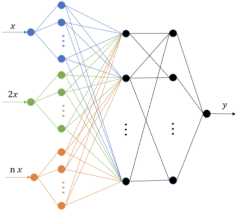
Idea: simultaneous training of frequency-selective subnetworks



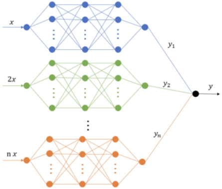
(b) sReLU



(a) sReLU



(a) MscaleDNN-1



(b) MscaleDNN-2

ANITI

Université Fédérale
Toulouse Midi-Pyrénées

Outline

Multigrid methods

Physics Informed Neural Networks (PINNs)

Multilevel PINNs (MPINNs)

Université Fédérale
Toulouse Midi-Pyrénées

Multilevel PINNs: formulation

Problem definition

$$\begin{aligned} D(z, u(z)) &= f(z), \; z \in \Omega, \\ u_{sol} &= u_H + u_h \end{aligned}$$

$$\mathbf{L}_h(\theta_h) = RMSE_{resh}(\theta_h) + RMSE_{datah}(\theta_h)$$

$$RMSE_{resh}(\theta_h) = \frac{\lambda^r}{N_h^r} ||D(\hat{u}_h + u_H) - f||^2$$

$$RMSE_{datah}(\theta_h) = \frac{\lambda^m}{N_h^m} ||\hat{u}_h + u_H - u||^2$$

With z_h the fine sampling

$$\mathbf{L}_H(\theta_H) = RMSE_{resH}(\theta_H) + RMSE_{dataH}(\theta_H)$$

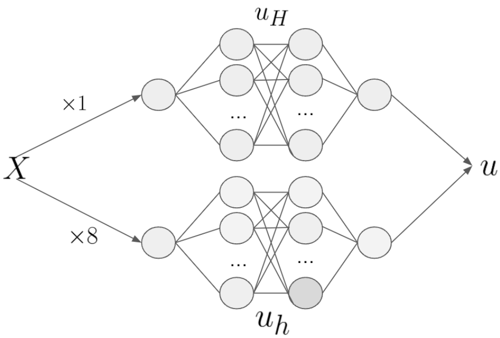
$$RMSE_{resH}(\theta_H) = \frac{\lambda^r}{N_H^r} ||D(\hat{u}_H + u_h) - f||^2$$

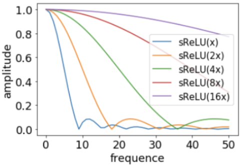
$$RMSE_{datah}(\theta_H) = \frac{\lambda^m}{N_H^m} ||\hat{u}_H + u_h - u||^2$$

With z_H the coarse sampling

Multilevel PINNs (0)

Also exploit frequency-selective subnetworks





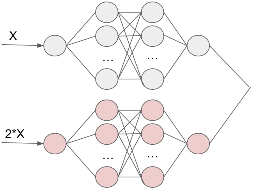
(a) sReLU

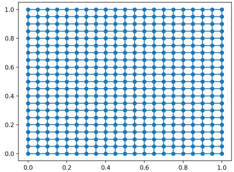
but. . .

Multilevel PINNs (1)

Algorithm 2-level training of PINNs

Freeze coarse-network parameters, unfreeze fine-network parameters
for i=1,2,... **do**
    Perform ν_1 epochs for the minimization of the fine problem
    Freeze fine-network parameters, unfreeze coarse-network parameters
    Perform ν_2 epochs for the minimization of the coarse problem
end for
Return : $u_H + u_h$

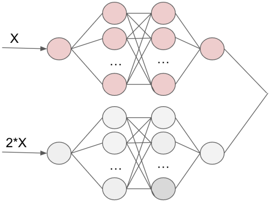


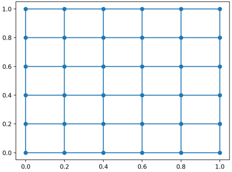


Multilevel PINNs (2)

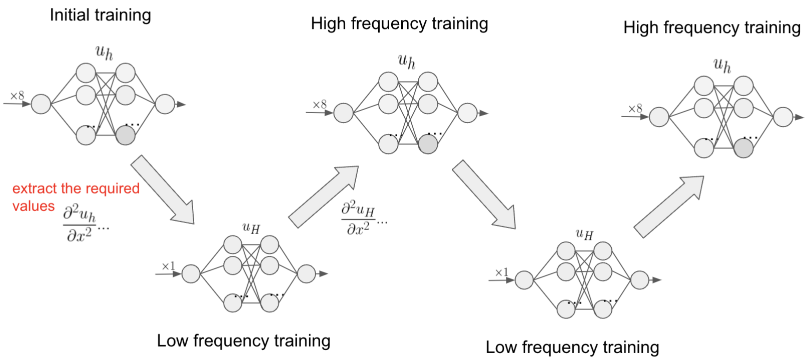
Algorithm 2-level training of PINNs

Freeze coarse-network parameters, unfreeze fine-network parameters
for i=1,2,... **do**
    Perform ν_1 epochs for the minimization of the fine problem
    Freeze fine-network parameters, unfreeze coarse-network parameters
    Perform ν_2 epochs for the minimization of the coarse problem
end for
Return : $u_H + u_h$



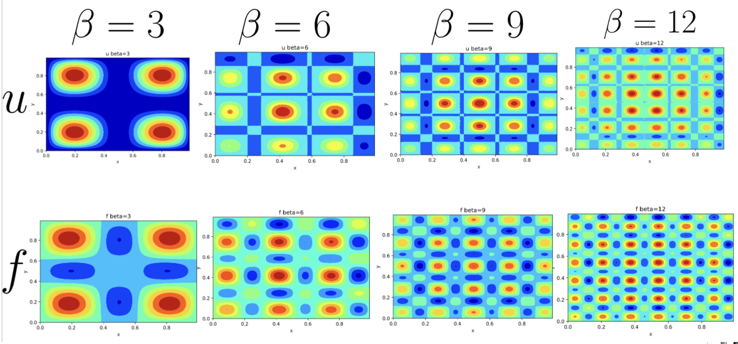


V-cycling between two levels



A simple Poisson problem

- $\Omega = [0,1] \times [0,1]$
- $\Delta u = f \ \forall x \in \Omega$
- $u = 0 \ \forall x \in \partial\Omega$
- $u(x,y) = (sin(\pi x) + sin(\beta \pi x)) * (sin(\pi y) + sin(\beta \pi y))$



Experimental settings

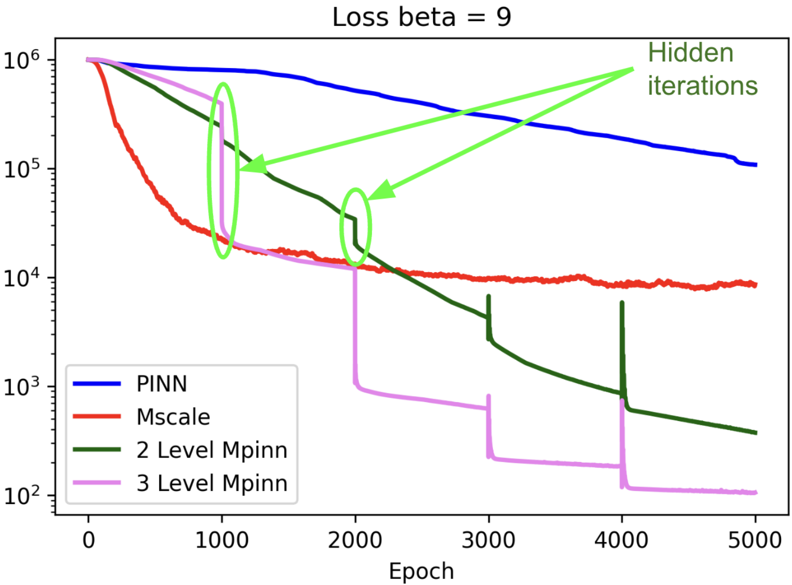
In what follows:

- The PINNs have two hidden layers of 300 neurons each.
- The Mscale have four subnetworks of two hidden layers of 150 neurons each, the input scaling used are 1,2,4 and 8.
- The two-level MPINN is composed of two networks of two hidden layers of 210 neurons each and trained in a V-cycle with 1 and 8 input scalings ($\nu_1 = \nu_2 = 1000$).
- The three-level MPINN is composed of three networks of two hidden layers of 150 neurons each and trained in a V-cycle with 1,4 and 8 input scalings ($\nu_1 = \nu_2 = \nu_3 = 1000$).
- The input of all networks is a regular grid sample of 80×80 points
- In all cases, we plot the median for ten random runs.

ANITI

Université Fédérale Toulouse Midi-Pyrénées

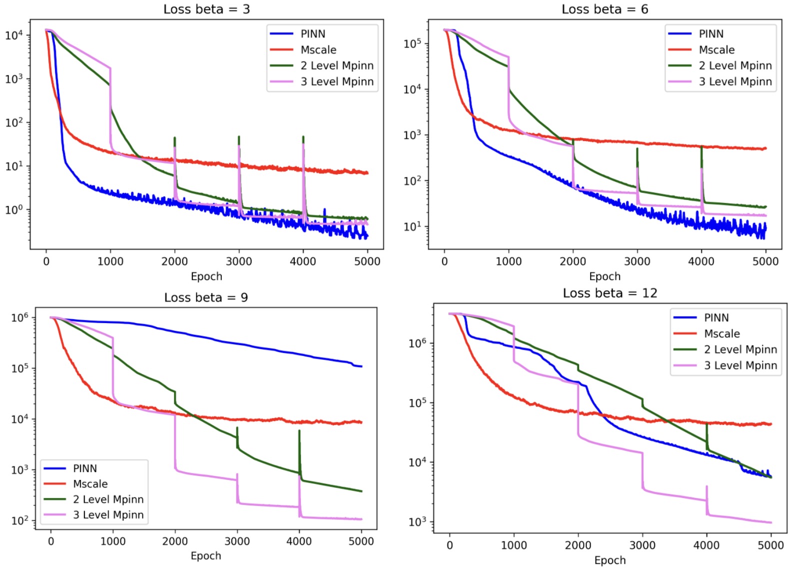
Experimental results



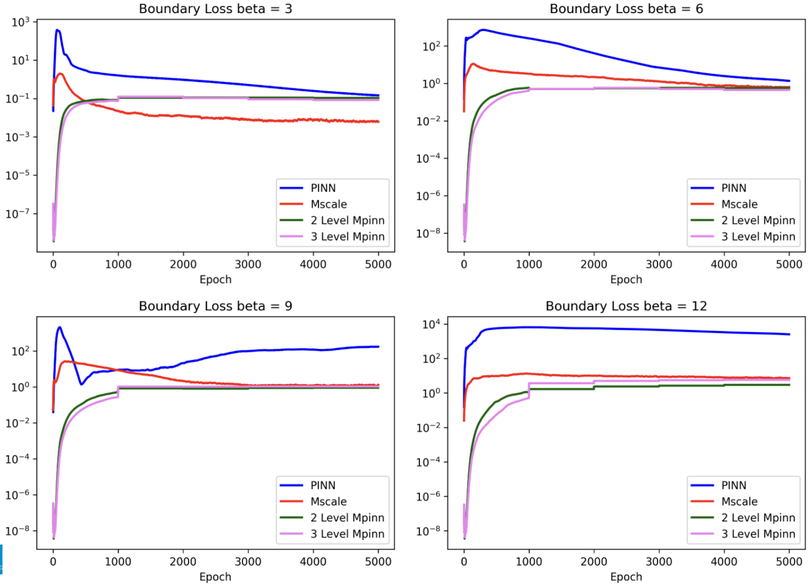
ANIT

Université Fédérale Toulouse Midi-Pyrénées

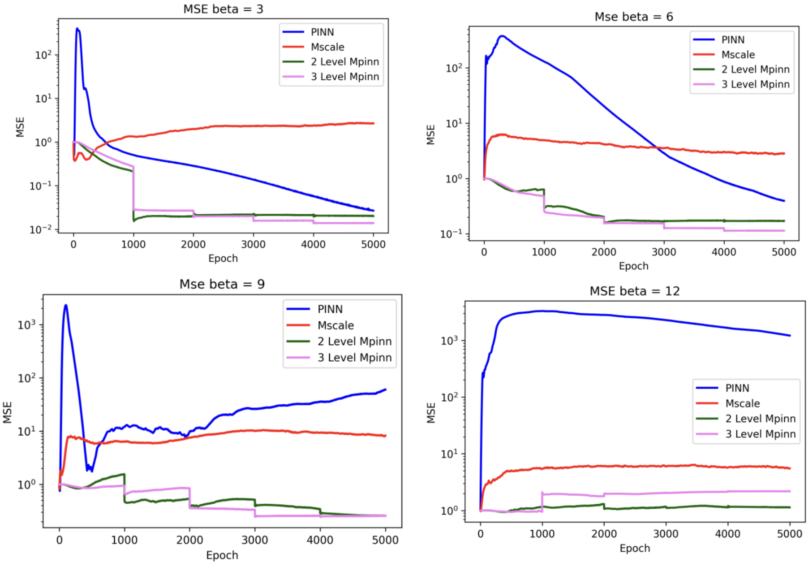
Varying β (the frequency content)



Convergence of error on boundary conditions

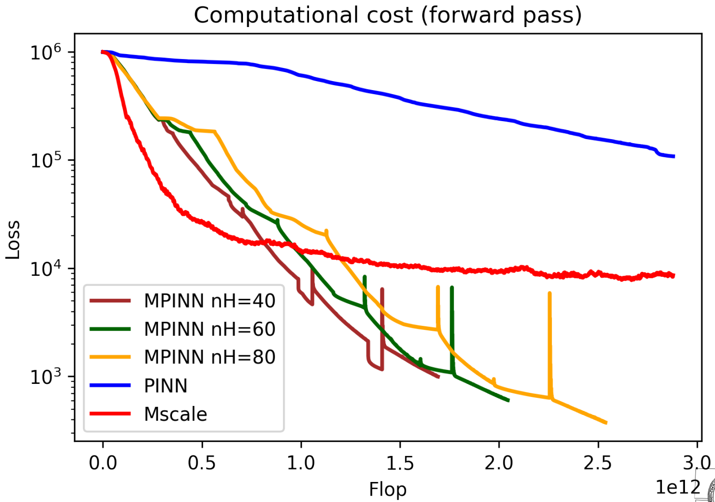


Convergence of MSE (extrapolation)



Computational cost for two levels. . .

. . . as a function of coarse grid size (nH)



Université Fédérale Toulouse Midi-Pyrénées

Conclusions

- We have presented a new multigrid-inspired training framework using recent advances in NN to efficiently solve PINN-type problems.
- We have proposed an algorithm which works without prior knowledge of frequency content and which is promising.
- We have demonstrated that exploiting spectral complementarity using our framework may bring significant computational benefits (faster convergence).

Perspectives

- Perform further extensive testing, including more complex problems.
- Pursue the sensitivity analysis for
  - the relative sizes of the grids,
  - the strategies for grid change.
- Investigate theoretical aspects:
  - convergence of the iterates from an optimization point of view,
  - convergence to the solution in functional space.
- Exploit the framework's versatility: extensions to other network types (e.g. deep-O-Net), as well as to other ways of targeting signal frequencies (Fourier feature mappings) or modelling complementarity.

A few references

- S. Gratton, A. Sartenaer, Ph. L. Toint. Recursive trust-region methods for multiscale nonlinear optimization, SIAM J. Opt., 19:414–444, 2008
- W. Briggs, V. Henson, S. McCormick. A Multigrid Tutorial, SIAM, 2000
- H. Calandra, S. Gratton, E. Riccietti, X. Vasseur. On high-order multilevel optimization strategies. SIAM J. Opt., 31.1: 307-330, 2021.
- S. Wang, H. Wang, P. Perdikaris. On the eigenvector bias of Fourier feature networks: From regression to solving multi-scale PDEs with physics-informed neural networks. CMAME, 384, 2021
- Z. Liu, W. Cai, Z. Xu. Multi-scale deep neural network (MscaleDNN) for solving Poisson-Boltzmann equation in complex domains. arXiv:2007.11207, 2020
- W. Cai, X. Li, L. Liu. A phase shift deep neural network for high frequency approximation and wave problems. SIAM Journal on Scientific Computing, 42(5), A3285-A3312, 2020.
- V. Mercier, S. Gratton, P. Boudier. A coarse space acceleration of deep-DDM, arxiv:2112.03732, 2022
- H. Calandra, S. Gratton, E. Riccietti, X. Vasseur. On a multilevel Levenberg-Marquardt method for the training of artificial neural networks and its application to the solution of partial differential equations, OMS, 2020