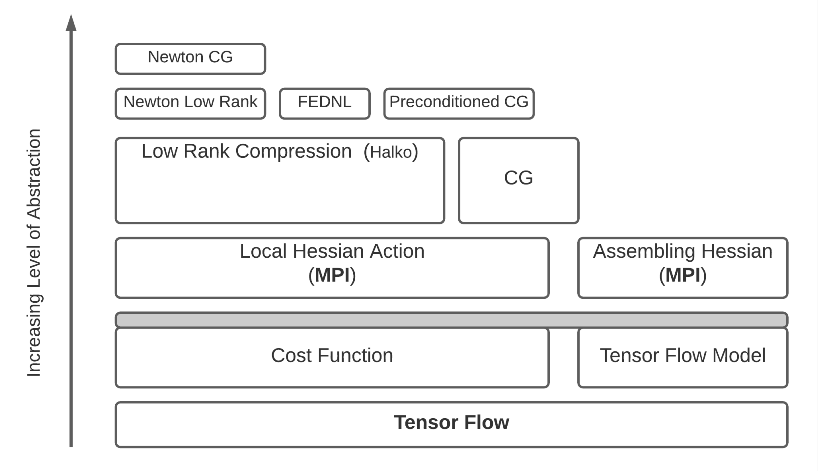
Solid,Dashed, Dotted and Dotted Dash Blue, Orange, Green and Red

Hessian Compression and PINNs

November 25, 2021

Developed Components

Developed Components



Hessian Compression

Algorithms presented in Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions by Halko, N. and Martinsson, P. G. were implemented to compress the Hessian.

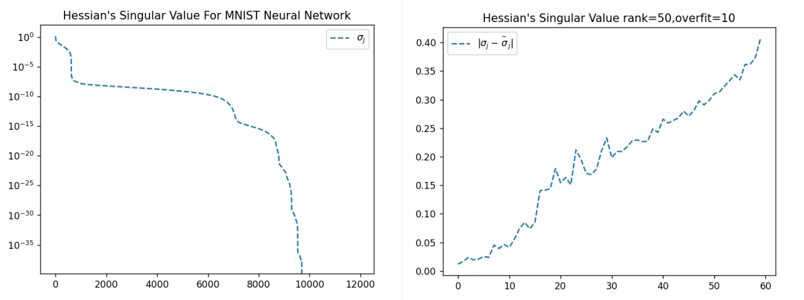


Figure 1: The graph refers to the Hessian of Neural Network trained on MNIST data set with a 15 node hidden layer with sigmoid activation function.

Convexity – Adult Salary Problem

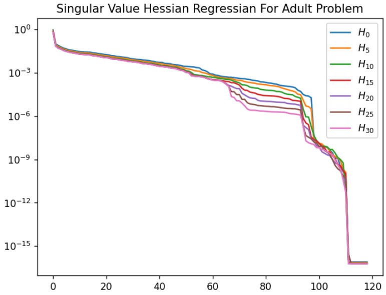
When working with convex energy functional one can use the components developed in order to minimize the energy of the problem.

In particular when working with convex energy second order method, such as the Newton method will converge to global minimums. In particular as a test problem we considered the adult slary problem with the following energy,

$$E(\vec{x}) = \frac{1}{m}\sum_{i=1}^{m} \log\left(1 + \exp\left(-b_j \vec{a}_j^{\,T}\vec{x}\right)\right)$$

$$\forall\, x \in \mathbb{R}^d$$

where d is the feature number and $\vec{a}_j$ are the data while b_j are the labels.

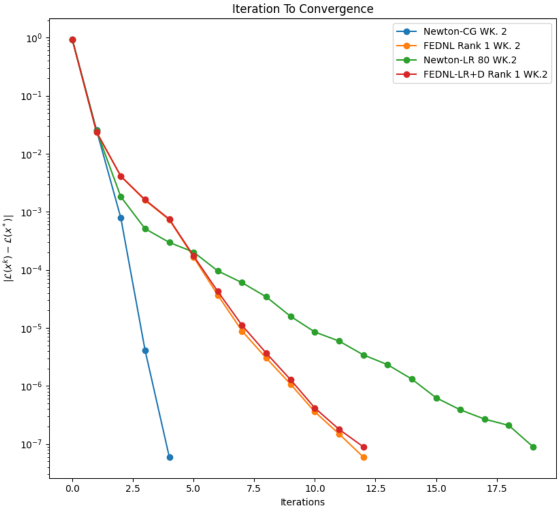


In particular the following Newton method was implemented,

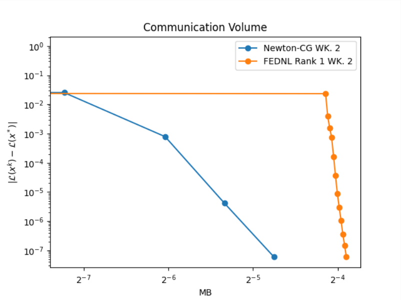
$$\vec{x}_{n+1} = \vec{x}_n - \gamma Hf(\vec{x}_n)^{-1}\nabla f(\vec{x}_n).$$

Method Comparison

Newton-CG, FEDNL, Newton-LR, FEDNL-LR+Diag



Newton-CG and FEDNL



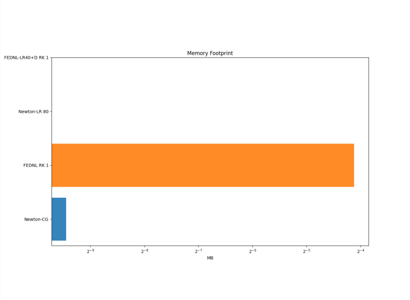


Figure 2: On the figure on the left one can observe the communication of FEDNL compared to Newton CG method.

Newton-LR and FEDNL

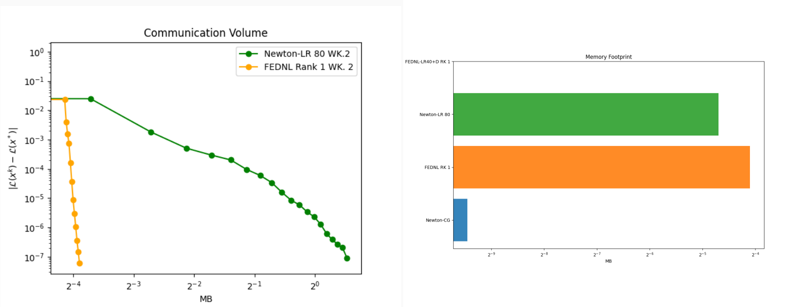


Figure 3: On the figure on the left one can observe the communication of FEDNL compared to Newton Low Rank method with rank compression 80. On the right figure memory footprint is compared.

FEDNL LR+Diagonal

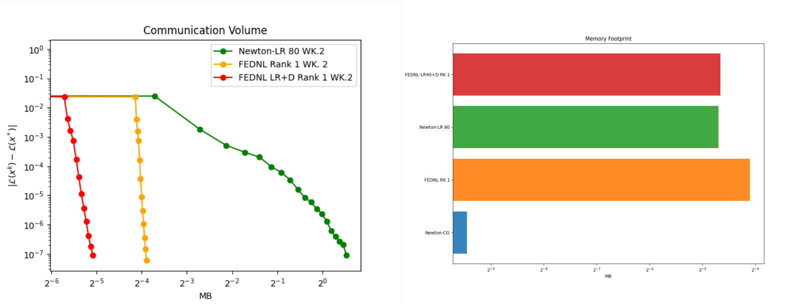
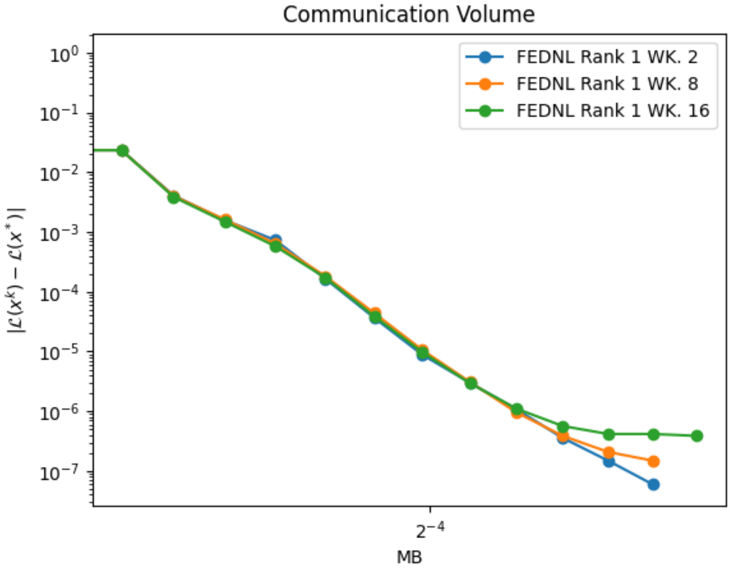


Figure 4: On the figure on the left one can observe the communication of FEDNL with Low Rank 40 + Diagonal compression compared to Newton Low Rank method with rank compression 80. On the right figure memory footprint is compared.



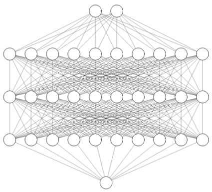
Communication Volume
FEDNL Rank 1 WK. 2
FEDNL Rank 1 WK. 8
FEDNL Rank 1 WK. 16
$|\mathcal{L}(x^k) - \mathcal{L}(x^*)|$
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
2^{-4}
MB

Neural Network And PDE

Physically Informed Neural Networks – PINNs

We would like to use Neural Network to solve partial differntial equations in 2D, in particular we would like to use the physically informed neural network.

The idea behind PINN is to construct a NN with the shape on the left and minimize the averaged least square residual on collocation points that are randomly sampled.



In particular the lost functional taken into consideration is,

$$\mathcal{L}(u,f) = \frac{1}{N}\sum_{i}^{N} |\Delta u(\vec{x_i}) - f(x_i)|^2 + \frac{1}{N}\sum_{i}^{N} |u(\vec{b_i})|^2 \quad (1)$$

$$\vec{x_i} \sim \mathcal{U}(\Omega) \qquad \vec{b_i} \sim \mathcal{U}(\partial\Omega) \qquad \Omega = [0,1]^2 \quad (2)$$

PINNs– Laplacian Eigenvalue

The minimization problem for the PINN require a long training in particular the figures below are computed using 50000 iteration of ADAM.

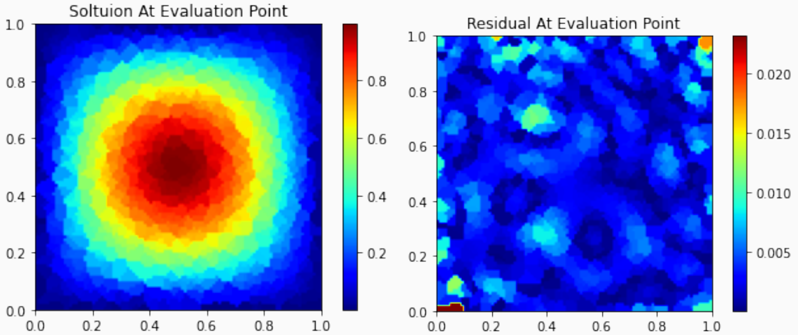
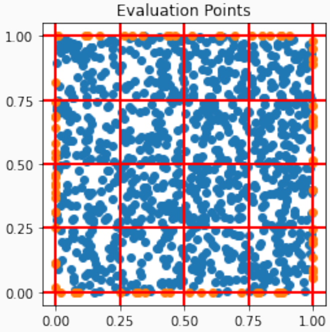


Figure 5: The figures above show the first eigenfunction of the Laplacian computed numerically and the residual both in the collocation points.

PINNs – Domain Decomposition

One advantage of PINNs is that is very easy to distribute the training of the problem in order to use a federated learning method such as FEDNL.



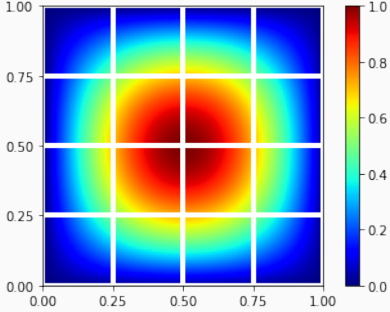


Figure 6: The figures show a domain decomposition, the idea is to distribute the lost function on each square among 16 different cores.

PINNs and GaLS

State Of The Art - Jinchao Xu

In recent work Jinchao Xu proved that shallow neural network with RELU activation function can retrieve the linear FEM space [J. Xu 2018]. In particular in the same paper the RELU DNN_1 is trained in order to minimize the Dirichlet energy functional,

$$J(u; f) = \frac{1}{2} \int_{\Omega} \nabla u \cdot \nabla u - f u dx. \tag{3}$$

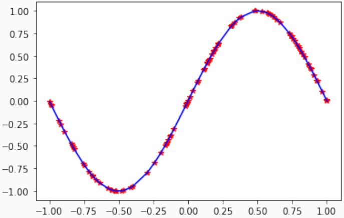
In particular this connection between RELU DNN1 functions allow to give a special meaning to the bias layer in the network, i.e. they are the mesh points of corresponding FEM tent functions. We would like to use similar techniques for physically informed neural networks, furthermore we aim to use recently developed idea for a priori analysis of stable neural network solutions [J. Xu 2021].

PINNs and GaLS

We now consider a shallow neural network with tanh activation funtion to solve the 1D Poisson problem.

We notice that the collocation residual that can be formulated as,

$$\mathcal{L}(u,f) = \frac{1}{N}\sum_{i}^{N} |\Delta u(\vec{x_i}) - f(x_i)|^2 + \frac{1}{N}\sum_{i}^{N} |u(\vec{b_i})|^2 \qquad (4)$$



can be see as the discrete version of the following least square energy functional,

$$J(u;f) = ||\Delta u - f||_0^2. \qquad (5)$$

In particular the discretisation is obtained using a mid point rule in this particular case.

PINNs and GaLS

Now is possible to prove that the DNN1 with tanh activation function retrieve a C1 conforming FEM space and therefore we can consider the following variation formulation of the above minimization problem,

$$a(u, v) = \int_\Omega \Delta u \Delta v = F(v) = \int_\Omega f \Delta v \tag{6}$$

which can be studied by means of a priori Rademacher complexity analysis [J. Xu 2021].

Future Work

Future Question

- Compressing the Hessian matrix using hierarchical methods.
- Use FEDNL to train PINNs on 2D and 3D Poisson problem.
- Use technique from Least Square Finite Elements to perform convergence analysis for PINNs and construct collocation residual.

Thank you !