SUPPLEMENTARY MATERIALS: UNDERSTANDING AND MITIGATING GRADIENT FLOW PATHOLOGIES IN PHYSICS-INFORMED NEURAL NETWORKS*

SIFAN WANG†, YUJUN TENG‡, AND PARIS PERDIKARIS‡

SM1. A bound for the gradients of PINNs boundary and residual loss functions for a one-dimensional Poisson problem. Recall that the loss function is given by

$$\begin{aligned}\mathcal{L}(\theta) &= \mathcal{L}_r(\theta) + \mathcal{L}_{u_b}(\theta) \\ &= \frac{1}{N_b}\sum_{i=1}^{N_b}[f_\theta(x_b^i) - h(x_b^i)]^2 + \frac{1}{N_r}\sum_{i=1}^{N_r}[\frac{\partial^2}{\partial x^2}f_\theta(x_r^i) - g(x_r^i)]^2.\end{aligned} \tag{SM1.1}$$

Here we fix $\theta \in \Theta$, where Θ denote all weights in a neural network. Then by assumptions, $\frac{\partial \mathcal{L}_{u_b}(\theta)}{\partial \theta}$ can be computed by

$$\begin{aligned}\left|\frac{\partial \mathcal{L}_{u_b}(\theta)}{\partial \theta}\right| &= \left|\frac{\partial}{\partial \theta}\left(\frac{1}{2}\sum_{i=1}^{2}\left(u_\theta\left(x_b^i\right) - h\left(x_b^i\right)\right)^2\right)\right| \\ &= \left|\sum_{i=1}^{2}\left(u_\theta\left(x_b^i\right) - h\left(x_b^i\right)\right)\frac{\partial u_\theta\left(x_b^i\right)}{\partial \theta}\right| \\ &= \left|\sum_{i=1}^{2}\left(u\left(x_b^i\right)\cdot\epsilon_\theta\left(x_b^i\right) - u\left(x_b^i\right)\right)u\left(x_b^i\right)\frac{\partial \epsilon_\theta\left(x_b^i\right)}{\partial \theta}\right| \\ &= \left|\sum_{i=1}^{2}u\left(x_b^i\right)\left(1 - \epsilon_\theta\left(x_b^i\right)\right)u\left(x_b^i\right)\frac{\partial \epsilon_\theta\left(x_b^i\right)}{\partial \theta}\right| \\ &\leqslant \left\|\frac{\partial \epsilon_\theta(x)}{\partial \theta}\right\|_{L^\infty}\cdot 2A\end{aligned}$$

Next, we may rewrite the $\mathcal{L}_r$ as

$$\mathcal{L}_r = \frac{1}{N_f}\sum_{i=1}^{N_f}\left|\frac{\partial^2 u_\theta}{\partial x^2}\left(x_f^i\right) - \frac{\partial^2 u}{\partial x^2}\left(x_f^i\right)\right|^2 \approx \int_0^1\left(\frac{\partial^2 u_\theta(x)}{\partial x^2} - \frac{\partial^2 u(x)}{\partial x^2}\right)^2 dx$$

*Supplementary material for SISC MS#M131804.
https://doi.org/10.1137/20M1318043
†Graduate Group in Applied Mathematics and Computational Science, University of Pennsylvania, Philadelphia, PA, 19104 (sifanw@sas.upenn.edu).
‡Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia, PA, 19104 (yjteng@seas.upenn.edu, pgp@seas.upenn.edu).

Then by integration by parts we have,

$$
\begin{aligned}
\frac{\partial \mathcal{L}_r}{\partial \theta} &= \frac{\partial}{\partial \theta} \int_0^1 \left(\frac{\partial^2 u_\theta(x)}{\partial x^2} - \frac{\partial^2 u(x)}{\partial x^2} \right)^2 dx \\
&= \int_0^1 \frac{\partial}{\partial \theta} \left(\frac{\partial^2 u_\theta(x)}{\partial x^2} - \frac{\partial^2 u(x)}{\partial x^2} \right)^2 dx \\
&= \int_0^1 2 \left(\frac{\partial^2 u_\theta(x)}{\partial x^2} - \frac{\partial^2 u(x)}{\partial x^2} \right) \frac{\partial}{\partial \theta} \left(\frac{\partial^2 u_\theta(x)}{\partial x^2} \right) dx \\
&= 2 \int_0^1 \left(\frac{\partial^2 u_\theta(x)}{\partial x^2} - \frac{\partial^2 u(x)}{\partial x^2} \right) \frac{\partial^2}{\partial x^2} \frac{\partial \left(u_\theta(x) \right)}{\partial \theta} \right) dx \\
&= 2 \left[\frac{\partial^2 u_\theta(x)}{\partial x \partial \theta} \left(\frac{\partial^2 u_\theta(x)}{\partial x^2} - \frac{\partial^2 u(x)}{\partial x^2} \right) \Big|_0^1 - \int_0^1 \frac{\partial^2 u_\theta(x)}{\partial x \partial \theta} \left(\frac{\partial^3 u_\theta(x)}{\partial x^3} - \frac{\partial^3 u(x)}{\partial x^3} \right) dx \right] \\
&= 2 \Bigg[\frac{\partial^2 u_\theta(x)}{\partial x \partial \theta} \left(\frac{\partial^2 u_\theta(x)}{\partial x^2} - \frac{\partial^2 u(x)}{\partial x^2} \right) \Big|_0^1 - \frac{\partial u_\theta(x)}{\partial \theta} \left(\frac{\partial^3 u_\theta(x)}{\partial x^3} - \frac{\partial^3 u(x)}{\partial x^3} \right) \Big|_0^1 \\
&+ \int_0^1 \frac{\partial u_\theta(x)}{\partial \theta} \left(\frac{\partial^4 u_\theta(x)}{\partial x^4} - \frac{\partial^4 u(x)}{\partial x^4} \right) dx \Bigg]
\end{aligned}
$$

Note that

$$
\begin{aligned}
\left| \frac{\partial^2 u_\theta(x)}{\partial x \partial \theta} \right| &= \left| \frac{\partial^2 u(x) \epsilon_\theta(x)}{\partial x \partial \theta} \right| = \left| \frac{\partial}{\partial \theta} \left(u'(x) \epsilon_\theta(x) + u(x) \epsilon'_\theta(x) \right) \right| \\
&= \left| u'(x) \frac{\partial \epsilon_\theta(x)}{\partial \theta} + u(x) \frac{\partial \epsilon'_\theta(x)}{\partial \theta} \right| \leq C \left\| \frac{\partial \epsilon_\theta(x)}{\partial \theta} \right\|_{L^\infty} + \left\| \frac{\partial \epsilon'_\theta(x)}{\partial \theta} \right\|_{L^\infty}
\end{aligned}
$$

And

$$
\begin{aligned}
\left| \frac{\partial^2 u_\theta(x)}{\partial x^2} - \frac{\partial^2 u(x)}{\partial x^2} \right| &= \left| \frac{\partial^2 u(x) \epsilon_\theta(x)}{\partial x^2} - \frac{\partial^2 u(x)}{\partial x^2} \right| \\
&= \left| u''(x) \epsilon_\theta(x) + 2u(x) \epsilon''_\theta(x) - u''(x) \right| \\
&= \left| u''(x) \left(\epsilon_\theta(x) - 1 \right) + 2u(x) \epsilon''_\theta(x) \right| \\
&\leq C^2 A + 2A
\end{aligned}
$$

So we have

$$
\left| \frac{\partial^2 u_\theta(x)}{\partial x \partial \theta} \left(\frac{\partial^2 u_\theta(x)}{\partial x^2} - \frac{\partial^2 u(x)}{\partial x^2} \right) \Big|_0^1 \right| \leq 2 \left(C \left\| \frac{\partial \epsilon_\theta(x)}{\partial \theta} \right\|_{L^\infty} + \left\| \frac{\partial \epsilon'_\theta(x)}{\partial \theta} \right\|_{L^\infty} \right) (C^2 A + 2A) \tag{SM1.2}
$$

$$
= O\left(C^3\right) \cdot A \cdot \left\| \frac{\partial \epsilon_\theta(x)}{\partial \theta} \right\|_{L^\infty} \tag{SM1.3}
$$

Similarly,

$$
\left| \frac{\partial u_\theta(x)}{\partial \theta} \left(\frac{\partial^3 u_\theta(x)}{\partial x^3} - \frac{\partial^3 u(x)}{\partial x^3} \right) \Big|_0^1 \right| = \left| \frac{\partial u_\theta(x)}{\partial \theta} \left(\frac{\partial^3 u(x) \epsilon_\theta(x)}{\partial x^3} - \frac{\partial^3 u(x)}{\partial x^3} \right) \Big|_0^1 \right| \tag{SM1.4}
$$

$$
\leq O\left(C^3\right) \cdot A \cdot \left\| \frac{\partial \epsilon_\theta(x)}{\partial \theta} \right\|_{L^\infty} \tag{SM1.5}
$$

Finally,

$$\left|\int_0^1 \frac{\partial u_\theta(x)}{\partial \theta}\left(\frac{\partial^4 u_\theta(x)}{\partial x^4} - \frac{\partial^4 u(x)}{\partial x^4}\right) dx\right| \leq O\left(C^4\right) \cdot A \cdot \left\|\frac{\partial \epsilon_\theta(x)}{\partial \theta}\right\|_{L^\infty} \tag{SM1.6}$$

Therefore, plugging all these together we obtain

$$\left|\frac{\partial \mathcal{L}_r}{\partial \theta}\right| \leq O\left(C^4\right) \cdot A \cdot \left\|\frac{\partial \epsilon_\theta(x)}{\partial \theta}\right\|_{L^\infty} \tag{SM1.7}$$

SM2. A proof of analyzing the difficulty in training PINNs with stiff gradient flow dynamics. By direct computation,

$$\mathcal{L}(\theta_{n+1}) - \mathcal{L}(\theta_n) = -\eta \nabla_\theta \mathcal{L}(\theta_n) \cdot \nabla_\theta \mathcal{L}(\theta_n) + \frac{1}{2}\eta^2 \nabla_\theta \mathcal{L}(\theta_n)^T \nabla_\theta^2 \mathcal{L}(\xi) \eta \nabla_\theta \mathcal{L}(\theta_n) \tag{SM2.1}$$

$$= -\eta \|\nabla_\theta \mathcal{L}(\theta_n)\|_2^2 + \frac{1}{2}\eta^2 \nabla_\theta \mathcal{L}(\theta_n)^T \nabla_\theta^2 \mathcal{L}(\xi) \nabla_\theta \mathcal{L}(\theta_n) \tag{SM2.2}$$

$$= -\eta \|\nabla_\theta \mathcal{L}(\theta_n)\|_2^2 + \frac{1}{2}\eta^2 \nabla_\theta \mathcal{L}(\theta_n)^T \left(\nabla_\theta^2 \mathcal{L}_r(\xi) + \nabla_\theta^2 \mathcal{L}_{u_b}(\xi)\right) \nabla_\theta \mathcal{L}(\theta_n) \tag{SM2.3}$$

Here, note that

$$\nabla_\theta \mathcal{L}(\theta_n)^T \nabla_\theta^2 \mathcal{L}(\xi) \nabla_\theta \mathcal{L}(\theta_n) = \|\nabla_\theta \mathcal{L}(\theta_n)\|_2^2 \frac{\nabla_\theta \mathcal{L}(\theta_n)^T}{\|\nabla_\theta \mathcal{L}(\theta_n)\|} \nabla_\theta^2 \mathcal{L}(\xi) \frac{\nabla_\theta \mathcal{L}(\theta_n)}{\|\nabla_\theta \mathcal{L}(\theta_n)\|} \tag{SM2.4}$$

$$= \|\nabla_\theta \mathcal{L}(\theta_n)\|_2^2 x^T Q^T \mathrm{diag}(\lambda_1, \lambda_2 \cdots \lambda_n) Q x \tag{SM2.5}$$

$$= \|\nabla_\theta \mathcal{L}(\theta_n)\|_2^2 y^T \mathrm{diag}(\lambda_1, \lambda_2 \ldots \lambda_M) y \tag{SM2.6}$$

$$= \|\nabla_\theta \mathcal{L}(\theta_n)\|_2^2 \sum_{i=1}^M \lambda_i y_i^2 \tag{SM2.7}$$

where $x = \frac{\nabla_\theta \mathcal{L}(\theta_n)}{\|\nabla_\theta \mathcal{L}(\theta_n)\|}$, Q is an orthogonal matrix diagonalizing $\nabla_\theta^2 \mathcal{L}(\xi)$ and $y = Qx$. And $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$ are eigenvalues of $\nabla_\theta^2 \mathcal{L}(\xi)$. Similarly, we have

$$\nabla_\theta \mathcal{L}(\theta_n)^T \nabla_\theta^2 \mathcal{L}_r(\xi) \nabla_\theta \mathcal{L}(\theta_n) = \|\nabla_\theta \mathcal{L}(\theta_n)\|_2^2 \sum_{i=1}^M \lambda_i^r y_i^2 \tag{SM2.8}$$

$$\nabla_\theta \mathcal{L}(\theta_n)^T \nabla_\theta^2 \mathcal{L}_{u_b}(\xi) \nabla_\theta \mathcal{L}(\theta_n) = \|\nabla_\theta \mathcal{L}(\theta_n)\|_2^2 \sum_{i=1}^M \lambda_i^{u_b} y_i^2 \tag{SM2.9}$$

where $\lambda_1^r \leq \lambda_2^r \leq \cdots \leq \lambda_N^r$ and $\lambda_1^{u_b} \leq \lambda_2^{u_b} \leq \cdots \leq \lambda_N^{u_b}$ are eigenvalues of $\nabla_\theta^2 \mathcal{L}_r$ and $\nabla_\theta^2 \mathcal{L}_{u_b}$ respectively. Thus, combining these together we get

$$\mathcal{L}(\theta_{n+1}) - \mathcal{L}(\theta_n) = \eta \|\nabla_\theta \mathcal{L}(\theta_n)\|_2^2 (-1 + \frac{1}{2}\eta \sum_{i=1}^N \lambda_i y_i^2) \tag{SM2.10}$$

$$\mathcal{L}_r(\theta_{n+1}) - \mathcal{L}_r(\theta_n) = \eta \|\nabla_\theta \mathcal{L}(\theta_n)\|_2^2 (-1 + \frac{1}{2}\eta \sum_{i=1}^N \lambda_i^r y_i^2) \tag{SM2.11}$$

$$\mathcal{L}_{u_b}(\theta_{n+1}) - \mathcal{L}_{u_b}(\theta_n) = \eta \|\nabla_\theta \mathcal{L}(\theta_n)\|_2^2 (-1 + \frac{1}{2}\eta \sum_{i=1}^N \lambda_i^{u_b} y_i^2) \tag{SM2.12}$$

SM3. Definitions of loss functions in Lid-Driven Cavity flow.

SM3.1. Velocity-pressure representation .

$$\mathcal{L}(\theta) = \mathcal{L}_{r_u}(\theta) + \mathcal{L}_{r_v}(\theta) + \mathcal{L}_{r_c}(\theta) + \mathcal{L}_{u_b}(\theta) + \mathcal{L}_{v_b}(\theta), \tag{SM3.1}$$

$$\mathcal{L}_{r_u}(\theta) = \frac{1}{N_r}\sum_{i=1}^{N_r}[r_\theta^u(x_r^i, y_r^i)]^2 \tag{SM3.2}$$

$$\mathcal{L}_{r_v}(\theta) = \frac{1}{N_r}\sum_{i=1}^{N_r}[r_\theta^v(x_r^i, y_r^i)]^2 \tag{SM3.3}$$

$$\mathcal{L}_{r_c}(\theta) = \frac{1}{N_r}\sum_{i=1}^{N_r}[r_\theta^c(x_r^i, y_r^i)]^2 \tag{SM3.4}$$

$$\mathcal{L}_{u_b}(\theta) = \frac{1}{N_b}\sum_{i=1}^{N_b}[u(x_b^i, y_b^i) - u_b^i]^2, \tag{SM3.5}$$

$$\mathcal{L}_{v_b}(\theta) = \frac{1}{N_b}\sum_{i=1}^{N_b}[v(x_b^i, y_b^i) - v_b^i]^2 \tag{SM3.6}$$

where $\{(x_r^i, y_r^i)\}_{i=1}^{N_r}$ is a set of collocation points in which we aim to minimize the PDE residual, while $\{(x_b^i, y_b^i), u_b^i\}_{i=1}^{N_b}$ and $\{(x_b^i, y_b^i), v_b^i\}_{i=1}^{N_b}$ denote the boundary data for the two velocity components at the domain boundaries Γ_0 and Γ_1, respectively.

SM4. Reference solution for a flow in a two-dimensional lid-driven cavity via a finite difference approximation. If we introduce the stream function ψ and vorticity ω, the Navier-Stokes equation can be written in the following form [SM1]:

$$\begin{cases} \dfrac{\partial \omega}{\partial t} + \dfrac{\partial \psi}{\partial y}\dfrac{\partial \omega}{\partial x} - \dfrac{\partial \psi}{\partial x}\dfrac{\partial \omega}{\partial y} = \nu \Delta \omega, \\ \Delta \psi = -\omega, \end{cases} \tag{SM4.1}$$

where

$$u = \frac{\partial \psi}{\partial y},\ v = -\frac{\partial \psi}{\partial x},\ \omega = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}. \tag{SM4.2}$$

The setup of the simulation is as follows. A set of points (x_i, y_j) is uniformly distributed in the domain $[0,1]\times[0,1]$, with $x_i = i/N,\ y_i = j/N,\ i,j = 0,1,\cdots,N$. The grid resolution h equals $1/N$. We denote $A_{i,j}$ as the value of physical variable A (velocity, pressure, etc.) at the point (x_i, y_j). All the spacial derivatives are treated with 2nd-order discretization scheme shown in SM4.3.

$$\begin{aligned} \frac{\partial A}{\partial x}|_{i,j} &\approx \frac{A_{i+1,j} - A_{i-1,j}}{2h}, \\ \frac{\partial A}{\partial y}|_{i,j} &\approx \frac{A_{i,j+1} - A_{i,j-1}}{2h}, \\ \Delta A|_{i,j} &\approx \frac{A_{i+1,j} + A_{i-1,j} + A_{i,j+1} + A_{i,j-1} - 4A_{i,j}}{h^2}. \end{aligned} \tag{SM4.3}$$

According to the boundary condition of velocity

$$(u,v) = \begin{cases} (1,0), & y=1, \\ (0,0), & \text{otherwise}, \end{cases} \tag{SM4.4}$$

we can derive the boundary condition of stream function

$$\psi = \begin{cases} h/2, & y=1, \\ 0, & \text{otherwise}, \end{cases} \tag{SM4.5}$$

under the assumption that the x-component of velocity u grows linearly from 0 to 1 between $(0, 1-1/N)$ and $(0,1)$, and between $(1, 1-1/N)$ and $(1,1)$.

The boundary condition of vorticity is derived from the Wood's formula [SM2]

$$\omega_0 = -\frac{1}{2}\omega_1 - \frac{3}{h^2}(\psi_1 - \psi_0) - \frac{3}{h}v_\tau - \frac{3}{2}\frac{\partial v_n}{\partial \tau} + \frac{h}{2}\frac{\partial^2 v_\tau}{\partial \tau^2}, \tag{SM4.6}$$

where (ψ_0, ω_0) is the local stream function and vorticity at a boundary point, (ψ_1, ω_1) is the stream function and vorticity at the adjacent point along the normal direction, (v_n, v_τ) is the normal and tangential component of velocity, and τ is the tangential direction.

The algorithm is composed of the following steps:

Step.1 Set the stream function ψ and vorticity ω at inner points to zero, and calculate the ψ and ω at the boundary using equation SM4.5 and SM4.6. Set time $t = 0$.

Step.2 Calculate the vorticity ω at inner points at the time $t + \Delta t$ with equation SM4.1(1), substituting $\partial\omega/\partial t$ with $(\omega(t+\Delta t) - \omega(t))/\Delta t$.

Step.3 Calculate the stream function ψ at inner points at the time $t + \Delta t$ with equation SM4.1(2) and boundary condition equation SM4.5.

Step.4 Calculate the vorticity ω on the boundary with equation SM4.6.

Step.5 Update the velocity (u, v) at the time $t+\Delta$ with equation SM4.2, and calculate the error between the velocity at the time t and $t + \Delta t$

$$\begin{aligned} \text{error}_u &= \frac{\max\limits_{i,j}\{u_{i,j}(t+\Delta t) - u_{i,j}(t)\}}{\Delta t}, \\ \text{error}_v &= \frac{\max\limits_{i,j}\{v_{i,j}(t+\Delta t) - v_{i,j}(t)\}}{\Delta t}. \end{aligned} \tag{SM4.7}$$

If $\max\{\text{error}_u, \text{error}_v\} < \varepsilon$, the flow has reached the steady state, and the computation terminates. Otherwise, set $t \leftarrow t + \Delta t$, and return to Step.2.

In our simulation, we set the number of grid points to $N = 128$, the time step to $\Delta t = 1 \times 10^3$, and the criterion for convergence to $\varepsilon = 1 \times 10^{-4}$.

REFERENCES

[SM1] U. Ghia, K. N. Ghia, and C. T. Shin, *High-Re solutions for incompressible flow using the Navier-Stokes equations and a multigrid method*, J. Comput. Phys., 48 (1982), pp. 387–411.

[SM2] L. C. Woods, *A note on the numerical solution of fourth order differential equations*, Aeronaut. Quart., 5 (1954), pp. 176–184.