A. Gradient issues with Physics Informed Neural Networks

PINNs, as any neural network, are not immune to standard neural network problems. Some of these are nonetheless exacerbated. The first of them was indicated by Raissi et al. (2020) - calculating a derivative of outputs with respect to inputs requires a full pass through the computational graph. This is especially serious for higher order derivatives which roughly double the size of the network with each additional order of derivative.

Another problem with PINNs, highlighted by Wang et al. (2020), are PINN gradient pathologies. The losses corresponding to the equations, containing the derivatives of the field functions, are calculated by multiple passes through the network. Thus, on average, the physics losses are an order of magnitude higher than the supervised losses, which, consequently, can cause the network to disregard any given data, producing a solution that obeys the governing physics but is, nonetheless, wrong. Wang et al. (2020) proposed a procedure which can successfully counter the phenomenon. Their solution is to scale the PINN data losses by the ratio of maximum absolute value of the PDE loss' gradients to the average absolute value of the data losses' gradients:

$$\text{scaling\_factor} = \frac{\max_j (|\nabla \mathcal{L}_{\mathcal{P},j}|)}{\text{avg}(|\nabla \mathcal{L}_{D,\cdot}|)}, \tag{28}$$

where each $\mathcal{L}_{D,\cdot}$ corresponds to a single data loss component (e.g. error between $\bar{u}$ measurements and PINN predictions) and each $\mathcal{L}_{\mathcal{P},\cdot}$ corresponds to a physics loss from a single equation of the pde system (e.g. mass conservation residual).

Our study uses an alternative solution. Instead of using gradient information to scale the losses, the values of the losses itself can be used to regularise the imbalance. The new scaling factor for a given data or boundary condition loss would be the ratio of that loss to the average value of PDE losses:

$$\text{scaling\_factor}_i = \frac{|\mathcal{L}_{D,i}|}{\text{avg}(|\mathcal{L}_{\mathcal{P},\cdot}|)} \tag{29}$$

This ensures that each of the data losses will be additionally penalised if it is too large when compared with the equation losses. Furthermore, to ensure stability, saturation function $\max(x,1)$ and momentum mechanism is added to the loss scaling, giving, for $k+1$'th iteration:

$$\text{scaling\_factor}_i^{k+1} = \alpha \cdot \text{scaling\_factor}_i^{k} + (1-\alpha) \cdot \max\left(\frac{|\mathcal{L}_{i;\text{DATA}}|}{\text{avg}(|\mathcal{L}_{\mathcal{P}}|)}, 1 \right) \tag{30}$$

where α is the momentum parameter that can be set to a value of around 0.9.

B. Distribution of data points for the PINN interpolations.

data resolution $\Delta x \times \Delta y$	bottom left corner point coordinates	upper right corner point coordinates
0.02×0.02	$(x, y) = (-1, -1.5)$	$(x, y) = (3, 1.5)$
0.05×0.05	$(x, y) = (-1, -1.5)$	$(x, y) = (3, 1.5)$
0.10×0.10	$(x, y) = (-1, -1.5)$	$(x, y) = (3, 1.5)$
0.25×0.25	$(x, y) = (-1, -1.5)$	$(x, y) = (3, 1.5)$
0.50×0.50	$(x, y) = (-1, -1.25)$	$(x, y) = (3, 1.25)$
0.60×0.60	$(x, y) = (-0.4, -1.2)$	$(x, y) = (3, 1.2)$
0.70×0.70	$(x, y) = (-0.75, -1.4)$	$(x, y) = (2.75, 1.4)$
1.00×1.00	$(x, y) = (-1, -1.5)$	$(x, y) = (3, 1.5)$

Table 6. *Data grid resolutions for interpolations along with coordinates of the bottom left and upper right corners of the data grid*

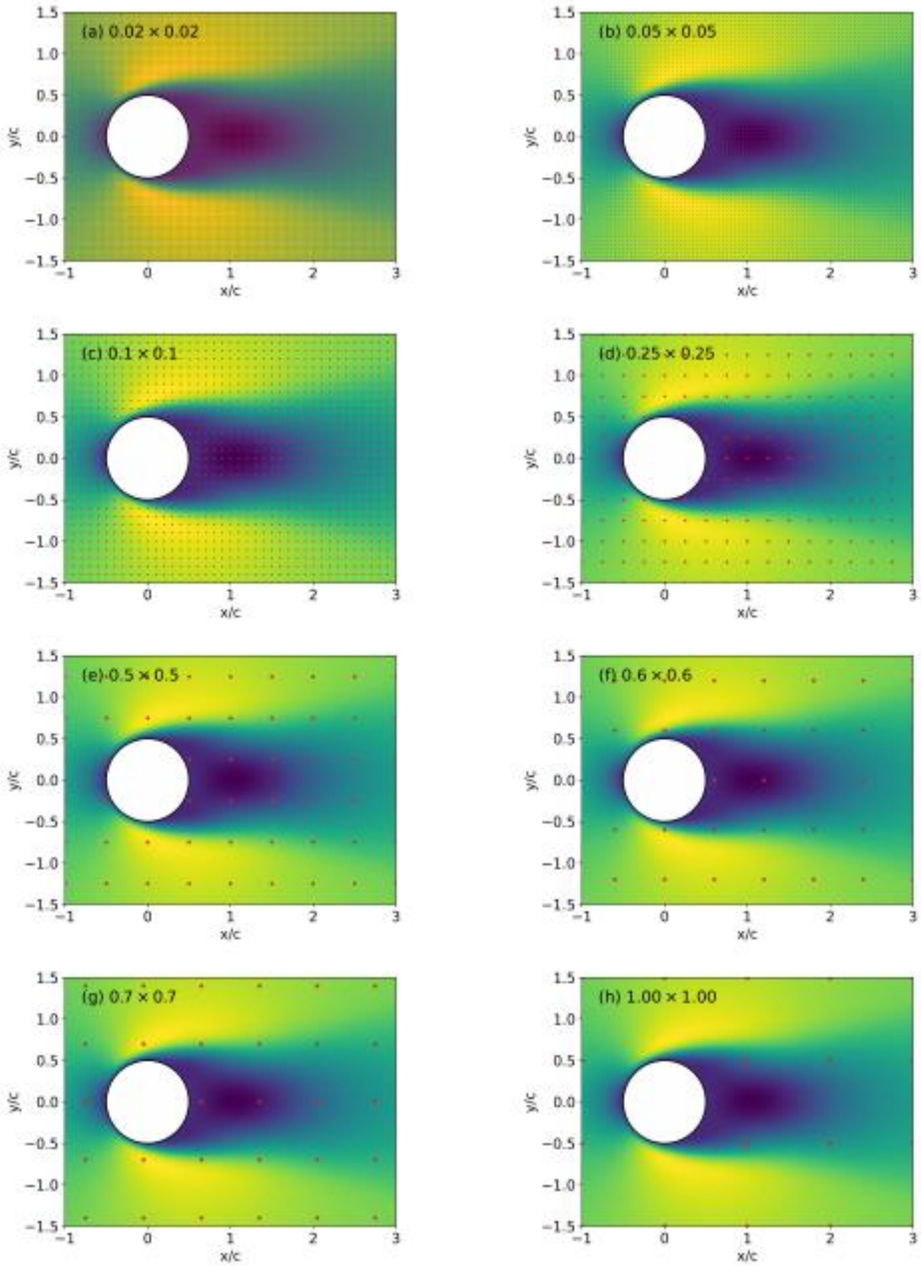

Figure 12. *Data grid points for various data grid densities plotted over true u velocity field.*

B.1. Practical range of resolutions for data from PIV experiments

The fine data spacing is 0.02 in both x and y directions. We will assume that during a PIV experiment that we need 16 pixels in each direction to calculate displacement of particles, and velocity. Then, for each single velocity vector that gives us $(3/0.02+1)\cdot 16=2416$ pixels to cover the 3 unit long height of the domain and $(4/0.02+1)\cdot 3216$ pixels for the 4 unit long width of the domain. This could be achieved with a 8 megapixel camera for the experimental PIV setup. Furthermore, using averaged data it is possible to combine PIV images from different experiments which again might be leveraged to increase the size of the input velocity data.