How does PDE order affect the convergence of PINNs?

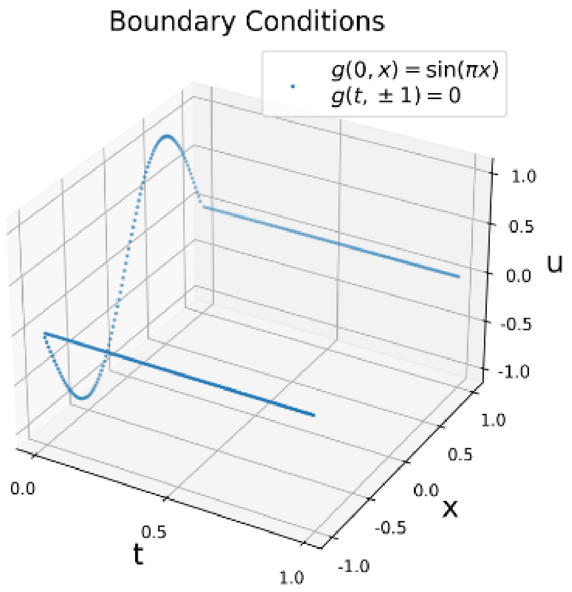
Changhoon Song, Yesom Park, Myungjoo Kang

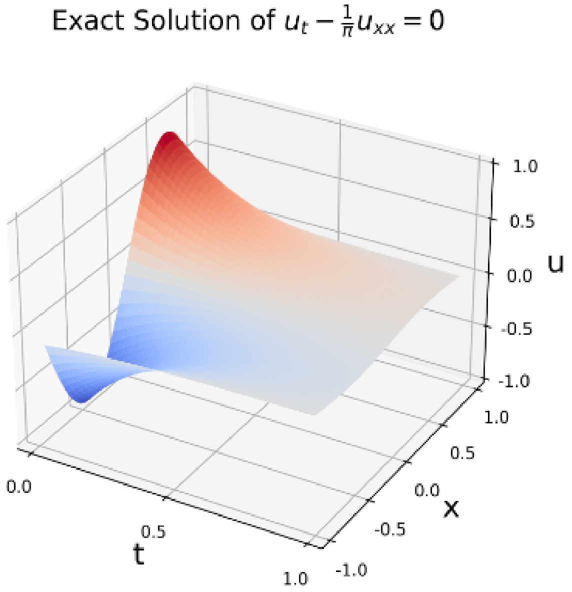
Partial Differential Equations

A partial differential equation(PDE) is an equation that computes a function between various partial derivatives of a multivariate function.

Examples:

- $\frac{\partial^2}{\partial x^2}u(x,y) + \frac{\partial^2}{\partial y^2}u(x,y) = 0.$
- $\frac{\partial}{\partial t}u(t,x) - \frac{\partial^2}{\partial x^2}u(t,x) = 0.$





Physics-Informed Neural Networks

A neural network u_θ is a solution of PDE if it satisfies

$$\begin{cases} \mathcal{N}\left[u_\theta, Du_\theta, D^2 u_\theta\right](\boldsymbol{x}) = f(\boldsymbol{x}), & \boldsymbol{x} \in \Omega, \\ u_\theta(\boldsymbol{x}) = g(\boldsymbol{x}), & \boldsymbol{x} \in \partial\Omega, \end{cases}$$

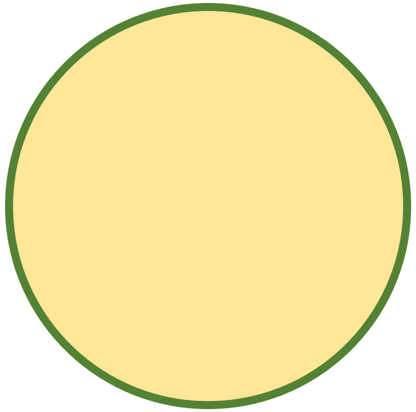
Physics-Informed Neural Networks (PINNs) [DPT94, RPK19]

PINNs learn a solution by minimizing the residual of the PDE:

$$\mathcal{L}(u_\theta) := \|\mathcal{N}[u_\theta] - f\|_{L^2(\Omega)} + \|u_\theta - g\|_{L^2(\partial\Omega)}.$$

Physics-Informed Neural Networks

Theoretical Setting



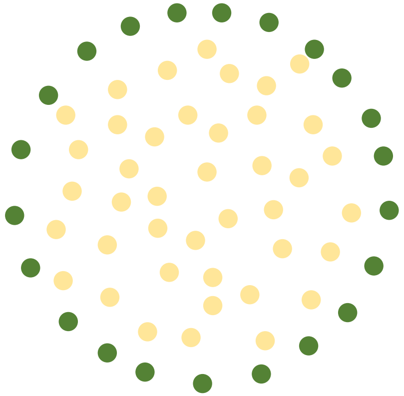
$\mathcal{N}[u_\theta] = f, \quad x \in \Omega$

$\mathcal{B}[u_\theta] = g, \quad x \in \partial\Omega$

$$\mathcal{L}(u_\theta) = \|\mathcal{N}[u_\theta] - f\|_{L^2(\Omega)} + \|\mathcal{B}[u_\theta] - g\|_{L^2(\partial\Omega)}$$

$$\theta = \arg\min_\theta \mathcal{L}(u_\theta)$$

Practical Setting



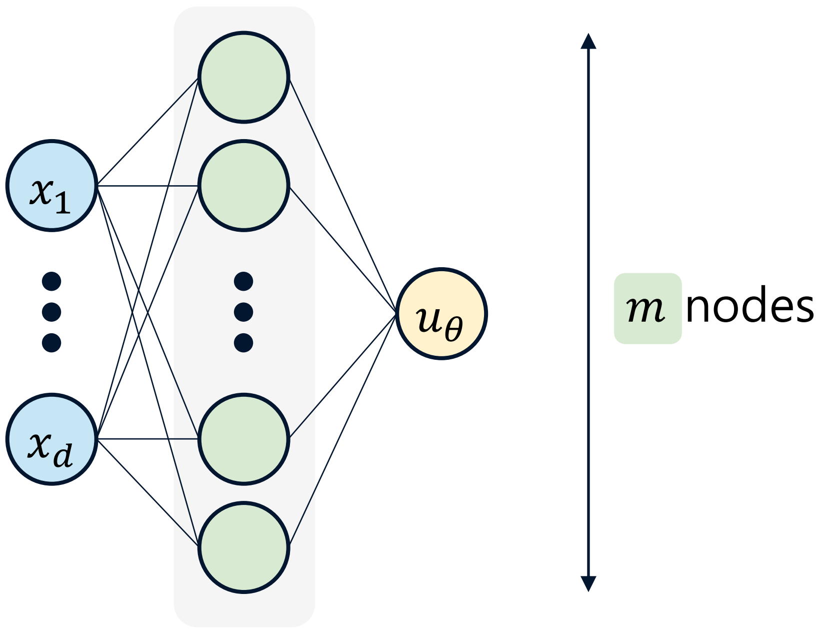
$\mathcal{N}[u_\theta](x_i) = f(x_i), \quad x_i \in \Omega$

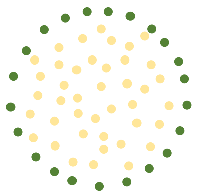
$\mathcal{B}[u_\theta](\tilde{x}_j) = g(\tilde{x}_j), \quad \tilde{x}_j \in \partial\Omega$

$$\mathcal{L}(u_\theta) = \sum_i \left(\mathcal{N}[u_\theta](x_i) - f(x_i)\right)^2 + \sum_j \left(\mathcal{B}[u_\theta](\tilde{x}_j) - g(\tilde{x}_j)\right)^2$$

$$\dot{\theta}(t) = -\nabla\mathcal{L}\left(u_{\theta(t)}\right)$$

Training Convergence of PINNs





$$\mathcal{N}[u_\theta](x_i) = f(x_i), \quad x_i \in \Omega$$

$$u_\theta(\tilde{x}_j) = g(\tilde{x}_j), \quad \tilde{x}_j \in \partial\Omega$$

$$\mathcal{N}[u] = \sum_{|\alpha| \le k} a_\alpha(\boldsymbol{x}) \frac{\partial^\alpha}{\partial \boldsymbol{x}^\alpha} u(\boldsymbol{x}), \qquad \mathcal{B}[u] = \sum_{|\alpha| \le 1} \tilde{a}_\alpha(\boldsymbol{x}) \frac{\partial^\alpha}{\partial \boldsymbol{x}^\alpha} u(\boldsymbol{x})$$

$$\mathcal{L}(u_\theta) = \sum_i \left(\mathcal{N}[u_\theta](x_i) - f(x_i)\right)^2 + \sum_j \left(\mathcal{B}[u_\theta](\tilde{x}_j) - g(\tilde{x}_j)\right)^2$$

$$u_\theta = \frac{1}{\sqrt{m}} \sum_{r=1}^{m} a_r\, \sigma(W_{ij} x_j + b_i)$$

$$\sigma(x) = \max\{0, x\}^p$$

Theorem (Brief)

$$m = \Omega\left(\log \frac{m}{\delta}\right)^{4p} \Longrightarrow P\left(\lim_{t\to\infty} \mathcal{L}(u(t)) = 0\right) \ge 1 - \delta.$$

Main result 1

Theorem (Special Case)

There exists a constant C, independent of d, k, and p, such that for any $\delta \ll 1$, if

$$m > C\binom{d+k}{d}^{14} p^{7k+4} 2^{6p} \left(\log \frac{md}{\delta}\right)^{4p}$$

then with probability of at least $1-\delta$ over the initialization, we have

$$\mathcal{L}_{PINN}(\boldsymbol{w}(t), \boldsymbol{v}(t)) \leq \exp(-\lambda_0 t)\, \mathcal{L}_{PINN}(\boldsymbol{w}(0), \boldsymbol{v}(0)), \ \forall t \geq 0.$$

Main result 1

Theorem (Special Case)

There exists a constant C, independent of d, k, and p, such that for any $\delta \ll 1$, if

PDE order

$$m > C\binom{d+k}{d}^{14} p^{7k+4} 2^{6p} \left(\log \frac{md}{\delta}\right)^{4p}$$

Width

then with probability of at least $1-\delta$ over the initialization, we have

$$\mathcal{L}_{PINN}\left(\boldsymbol{w}(t), \boldsymbol{v}(t)\right) \leq \exp(-\lambda_0 t)\, \mathcal{L}_{PINN}\left(\boldsymbol{w}(0), \boldsymbol{v}(0)\right), \ \forall t \geq 0.$$

Loss at time t

Initial loss

- Higher k and p requires exponentially wide width.
- $p = k + 1$ is optimal order for RePU, since $p \geq k + 1$.

Main result 1

Theorem (Special Case)

There exists a constant C, independent of d, k, and p, such that for any $\delta \ll 1$, if

$$m > C\binom{d+k}{d}^{14} p^{7k+4} 2^{6p} \left(\log \frac{md}{\delta}\right)^{4p}$$

Width

RePU power

then with probability of at least $1-\delta$ over the initialization, we have

$$\mathcal{L}_{PINN}(\boldsymbol{w}(t), \boldsymbol{v}(t)) \le \exp(-\lambda_0 t)\, \mathcal{L}_{PINN}(\boldsymbol{w}(0), \boldsymbol{v}(0)), \quad \forall t \ge 0.$$

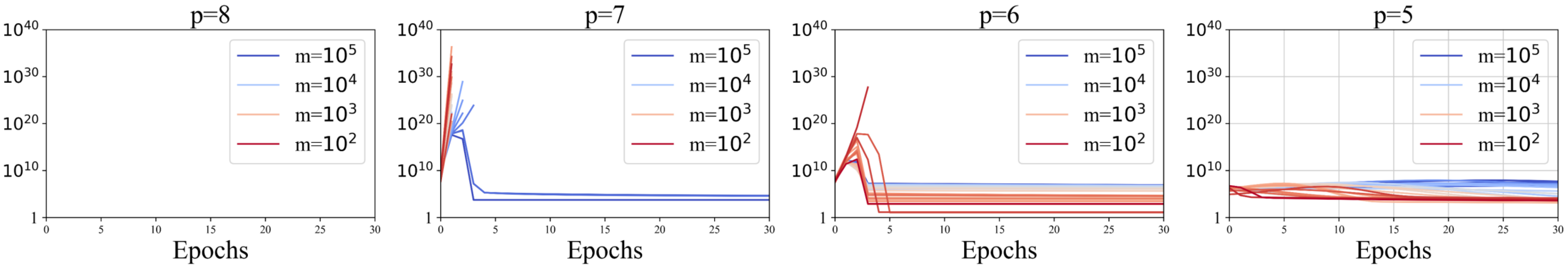
Loss at time t

Initial loss

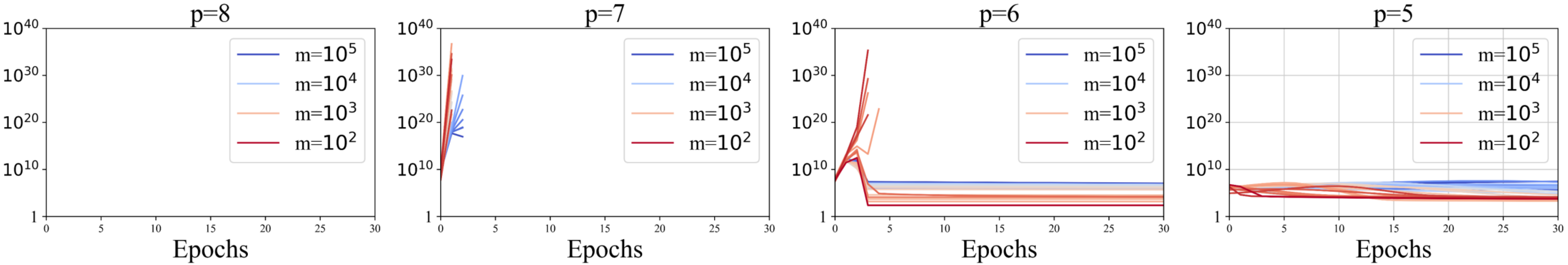
- Higher k and p requires exponentially wide width.
- $p = k + 1$ is optimal order for RePU, since $p \ge k + 1$.

Experiment 1

Harmonic equation: $u_{xx} + u_{yy} = f_1$

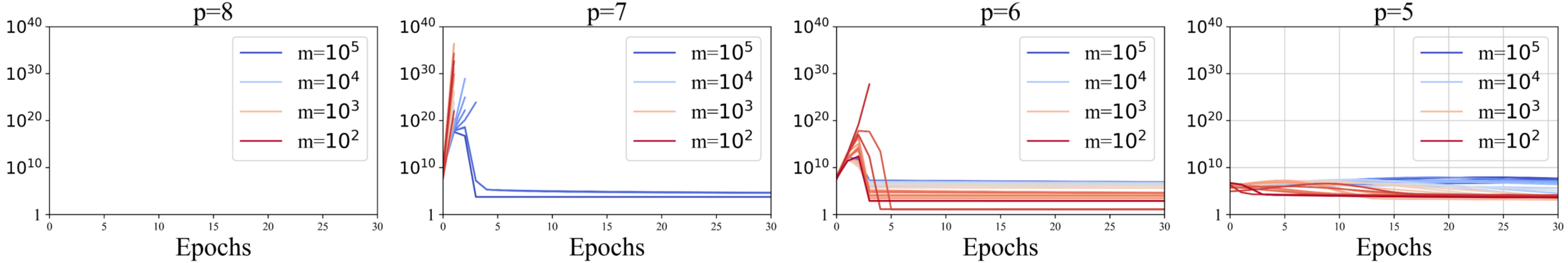


Biharmonic equation: $u_{xxxx} + 2u_{xxyy} + u_{yyyy} = f_2$



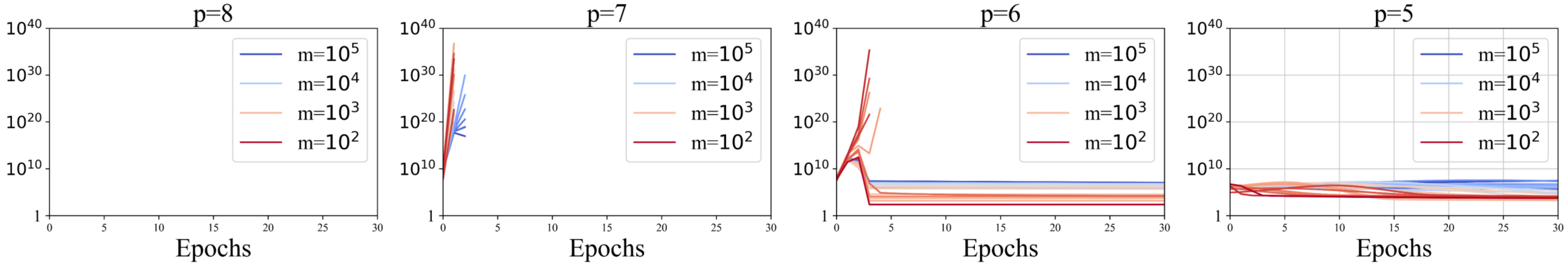
Experiment 1

Harmonic equation: $u_{xx} + u_{yy} = f_1$



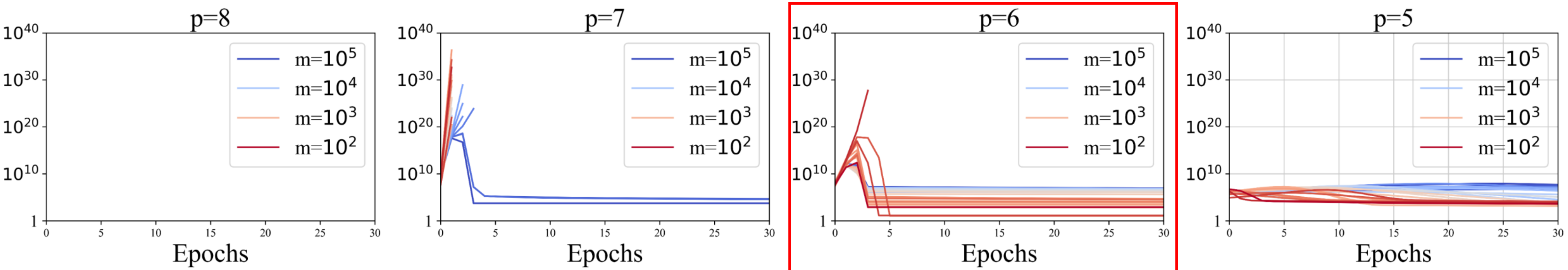
Larger p requires larger m

Biharmonic equation: $u_{xxxx} + 2u_{xxyy} + u_{yyyy} = f_2$

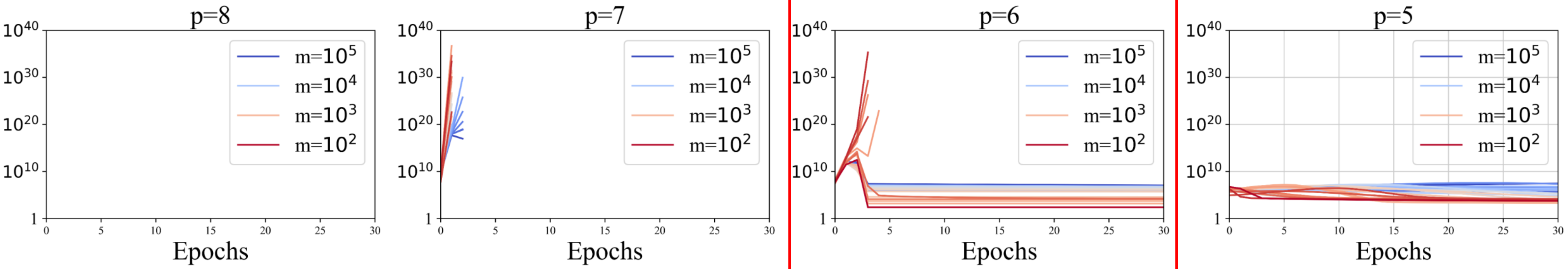


Experiment 1

Harmonic equation: $u_{xx} + u_{yy} = f_1$



Biharmonic equation: $u_{xxxx} + 2u_{xxyy} + u_{yyyy} = f_2$



Larger k requires larger m

Variable Splitting

$$\Delta u = f$$

primary variable

Auxiliary variable

$$\begin{cases} \nabla \cdot V = f \\ V = \nabla u \end{cases}$$

Higher-order PDEs

$$\begin{cases} \mathcal{N}[u](\boldsymbol{x}) = f(\boldsymbol{x}), \\ \mathcal{B}[u](\boldsymbol{x}) = g(\boldsymbol{x}), \end{cases}$$

$$\mathcal{N}[u] = \sum_{|\alpha| \le k} a_\alpha \frac{\partial^\alpha}{\partial \boldsymbol{x}^\alpha} u$$

System of lower-order PDEs

$$\begin{cases} \hat{\mathcal{N}}[\phi_0, \cdots, \phi_L](\boldsymbol{x}) = f(\boldsymbol{x}), \\ \frac{\partial^\beta}{\partial \boldsymbol{x}^\beta} (\phi_{\ell-1})_\alpha (\boldsymbol{x}) = (\phi_\ell)_{\alpha+\beta} (\boldsymbol{x}) \\ \mathcal{B}[\phi_0](\boldsymbol{x}) = g, \end{cases}$$

$$\mathcal{N}[u] = \sum_{\ell} \sum_{|\alpha| \le \xi_\ell} \sum_{|\beta| \le \Delta \xi_{\ell+1}} \hat{a}_{\ell,\alpha,\beta} \frac{\partial^{\Delta \xi_{\ell+1}}}{\partial \boldsymbol{x}^\beta} (\phi_\ell)_\alpha$$

Main result 2

Theorem (General Case)

There exists a constant C, independent of d, k, $|\xi|$, and p, such that for any $\delta \ll 1$, if

$|\xi|$: maximal order in system of PDEs

$$m > C \binom{d+k}{d}^6 \binom{d+|\xi|}{d}^8 p^{7|\xi|+4} 2^{6p} \left(\log \frac{md}{\delta}\right)^{4p},$$

then with probability of at least $1-\delta$ over the initialization, we have

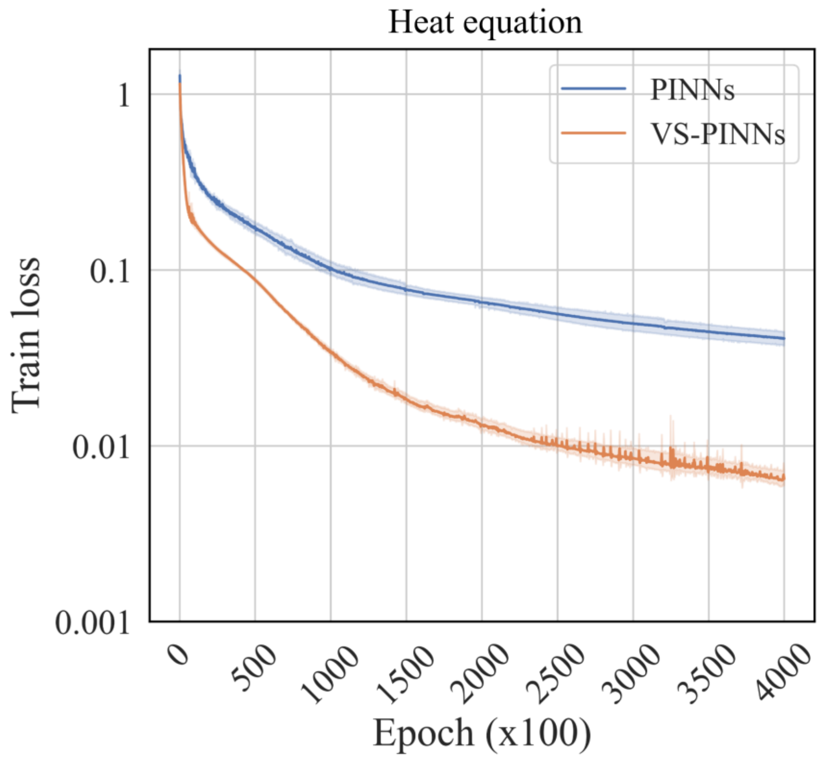
$$\mathcal{L}_{PINN}^{VS}(\mathbf{w}(t), \mathbf{v}(t)) \le \exp(-\lambda_0 t)\, \mathcal{L}_{PINN}^{VS}(\mathbf{w}(0), \mathbf{v}(0)), \ \forall t \ge 0.$$

- Lower $|\xi|$ reduces width requirement.
- $p = |\xi| + 1$ is optimal order for RePU.

Experiment 2

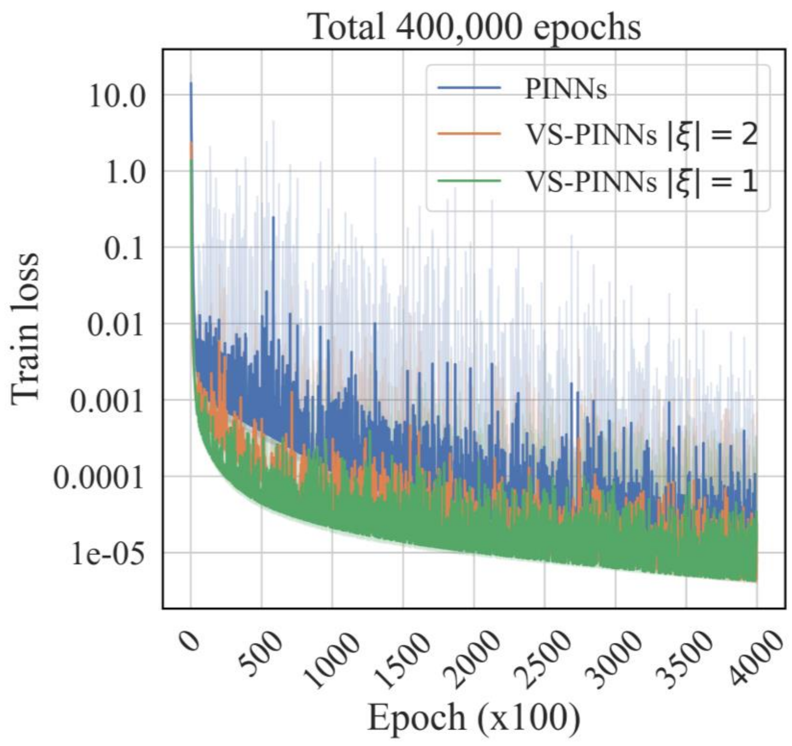
Heat equation (GD)

$$\begin{cases} u_t = u_{xx} \\ u(t,-1) = u(t,1) = 0 \\ u(0,x) = \sin(\pi x) \end{cases}$$



Elastic beam equation (Adam)

$$\begin{cases} u_t + u_{xxxx} = 0 \\ u(t,0) = u(t,\pi) = u_{xx}(t,0) = u_{xx}(t,\pi) = 0 \\ u(0,x) = 2\sin(x) \end{cases}$$



Summary

- The PINNs converge to global minimizer, provided enough network size.
- The higher the PDE order, the larger the network should be.
- Re-formulating high-order PDE as system of lower-order PDEs enhance the convergence condition.

Thank you