How does PDE order affect the convergence of PINNs?

Changhoon Song, Yesom Park, Myungjoo Kang

24.11.06

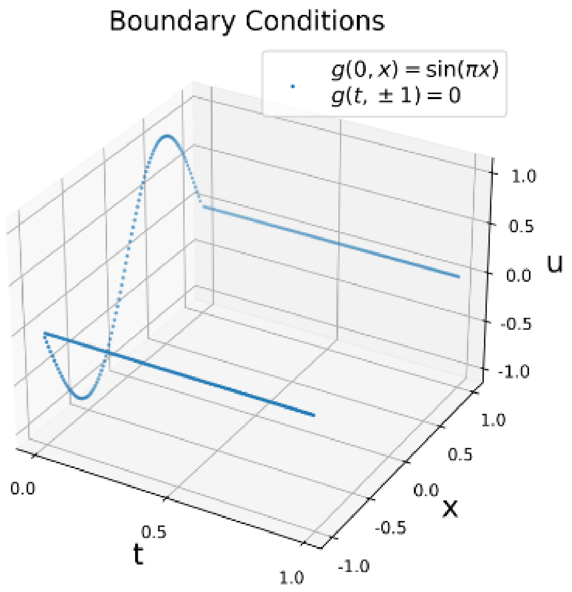
Contents

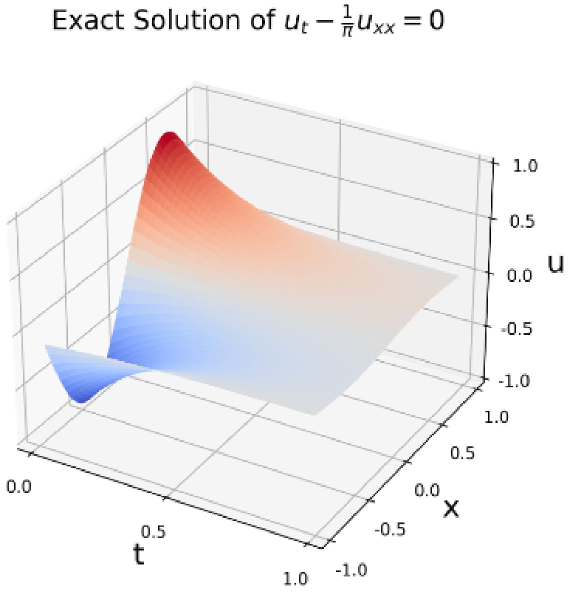
Partial Differential Equations

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

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

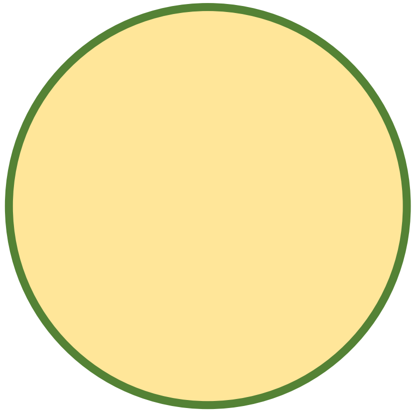
Physics-Informed Neural Networks (PINNs) [DPT94, RPK19]

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



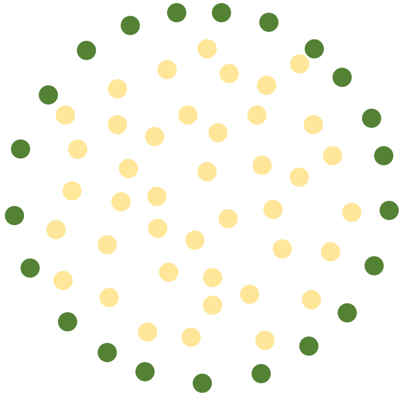
$\mathcal{N}[u_\theta] = f, \quad x \in \Omega$

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

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

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

Practical Setting



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

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

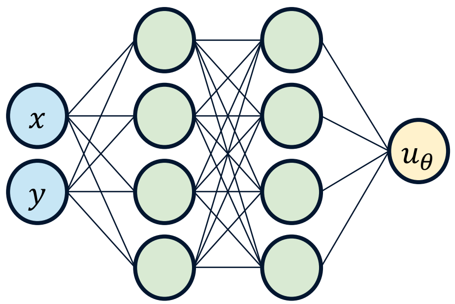
$$\mathcal{L}(u_\theta) = \sum_i \left(\mathcal{N}[u_\theta](x_i) - f(x_i)\right)^2 + \sum_j \left(\mathcal{B}[u_\theta](\tilde{x}_j) - g(\tilde{x}_j)\right)^2$$

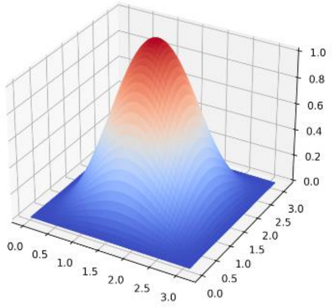
$$\theta(t+1) = \theta(t) - \eta \nabla \mathcal{L}\left(u_{\theta(t)}\right)$$

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

Convergence of PINNs

Network ≈ Exact sol.





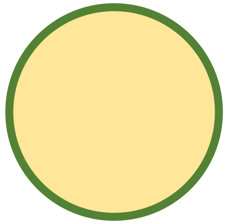
Minimize residual ≈ Minimize error

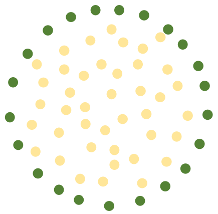
$$\mathcal{L}(u_\theta) = \|\mathcal{N}[u_\theta] - f\|_{L^2(\Omega)} + \|\mathcal{B}[u_\theta] - g\|_{L^2(\partial\Omega)}$$

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

$$\|u_\theta - u^*\| \leq C\mathcal{L}(u_\theta)$$

Expected residual ≈ Sampled residual





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

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

Training result ≈ Loss minimizer

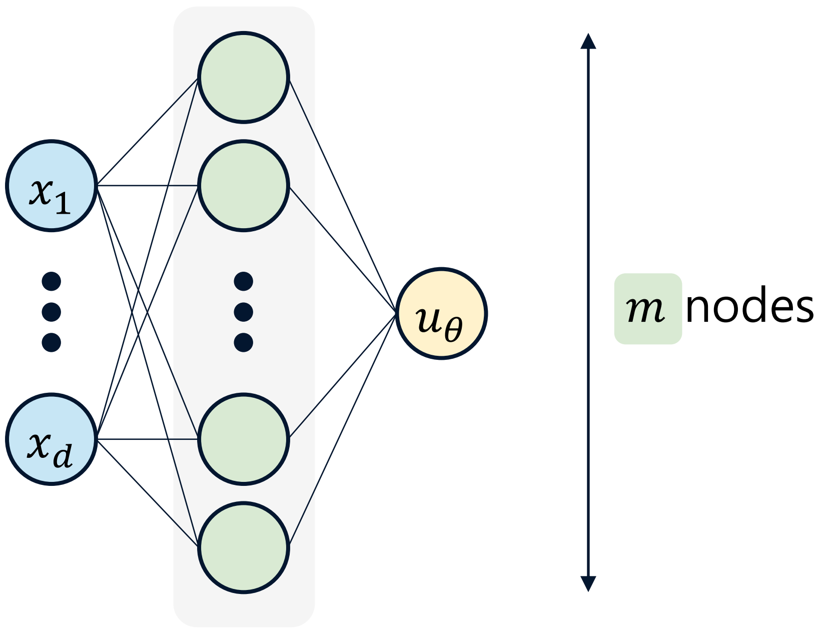
$$\theta(t+1) = \theta(t) - \eta\nabla\mathcal{L}\left(u_{\theta(t)}\right)$$

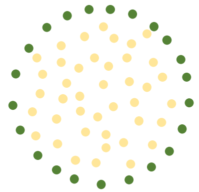
$$\dot{\theta}(t) = -\nabla\mathcal{L}\left(u_{\theta(t)}\right)$$

$$\lim_{t\to\infty} \theta(t)$$

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

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

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

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

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

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

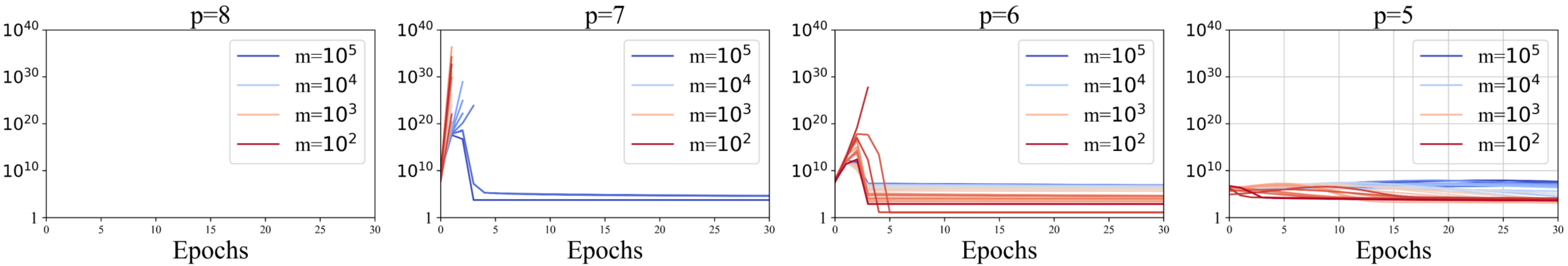
$$\mathcal{L}_{PINN}(\boldsymbol{w}(t), \boldsymbol{v}(t)) \leq \exp(-\lambda_0 t) \mathcal{L}_{PINN}(\boldsymbol{w}(0), \boldsymbol{v}(0)), \quad \forall t \geq 0.$$

Loss at time t — Initial loss

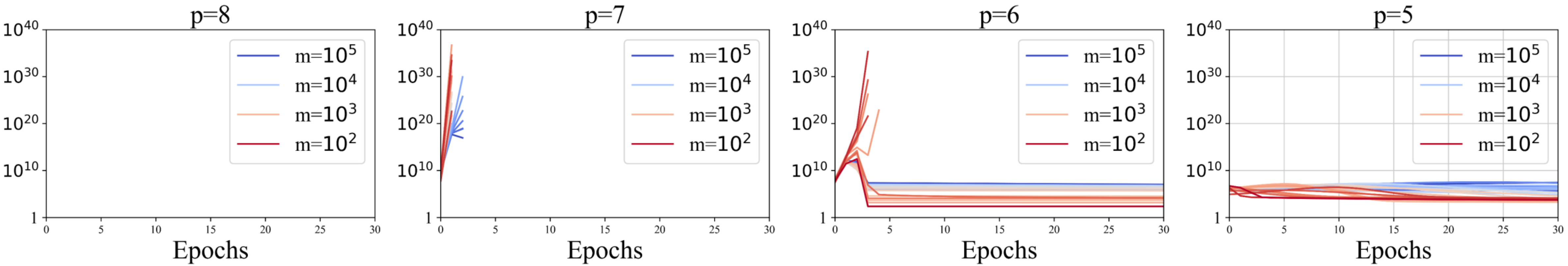
- Higher k and p requires exponentially wide width.
- $p = k + 1$ is optimal order for RePU, since $p \geq k + 1$.

Experiment 1

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



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



Variable Splitting

Higher-order PDEs

System of lower-order PDEs

$$\Delta u = f$$

primary variable

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

Auxiliary variable

$$u_t - u_{xx} = f$$

$$\begin{cases} u_t - v_x = f \\ v = u_x \end{cases}$$

Auxiliary variable

Variable Splitting

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

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

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

$$\boldsymbol{0 = \xi_0 \le \xi_1 \le \cdots \le \xi_{L+1} = k}$$

$$\boldsymbol{\Delta\xi_\ell = \xi_{\ell+1} - \xi_\ell}$$

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

$|\xi| = \max \Delta\xi_\ell$

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

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

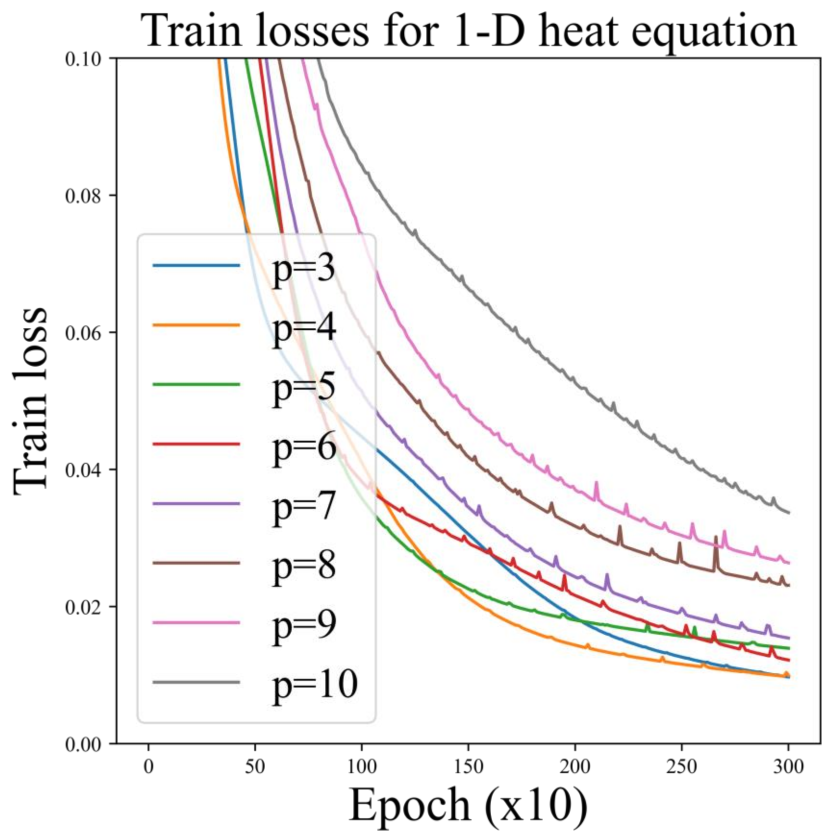
$$\mathcal{L}_{PINN}^{VS}(\mathbf{w}(t), \mathbf{v}(t)) \leq \exp(-\lambda_0 t) \mathcal{L}_{PINN}^{VS}(\mathbf{w}(0), \mathbf{v}(0)), \ \forall t \geq 0.$$

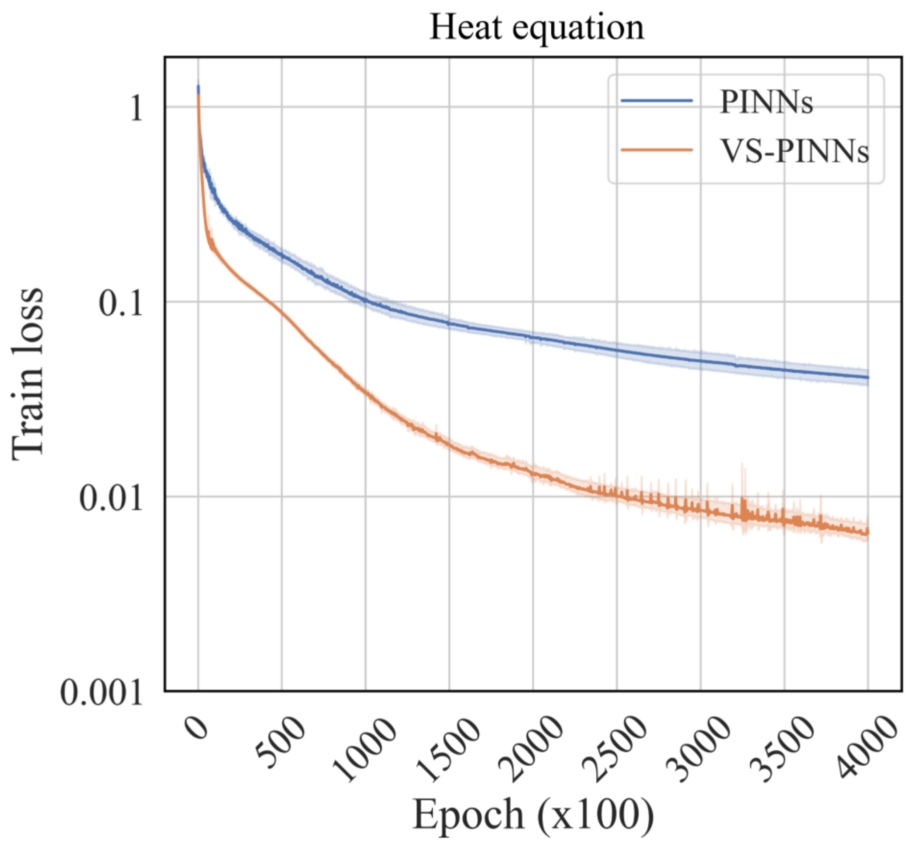
- Lower $|\xi|$ reduces width requirement.
- $p = |\xi| + 1$ is optimal order for RePU.

Experiment 2

Heat equation

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

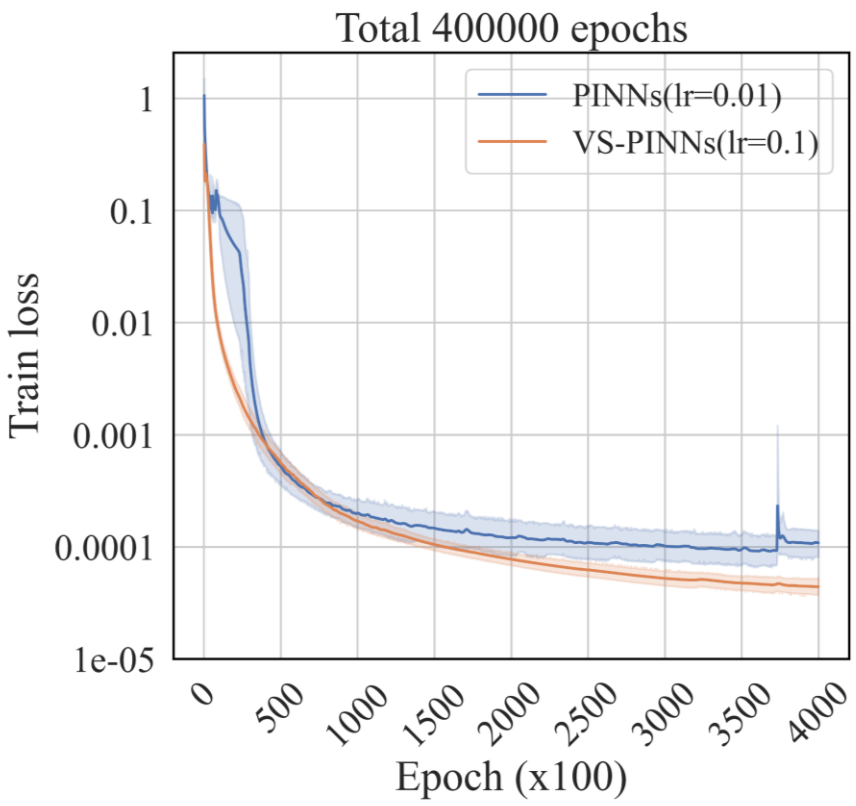


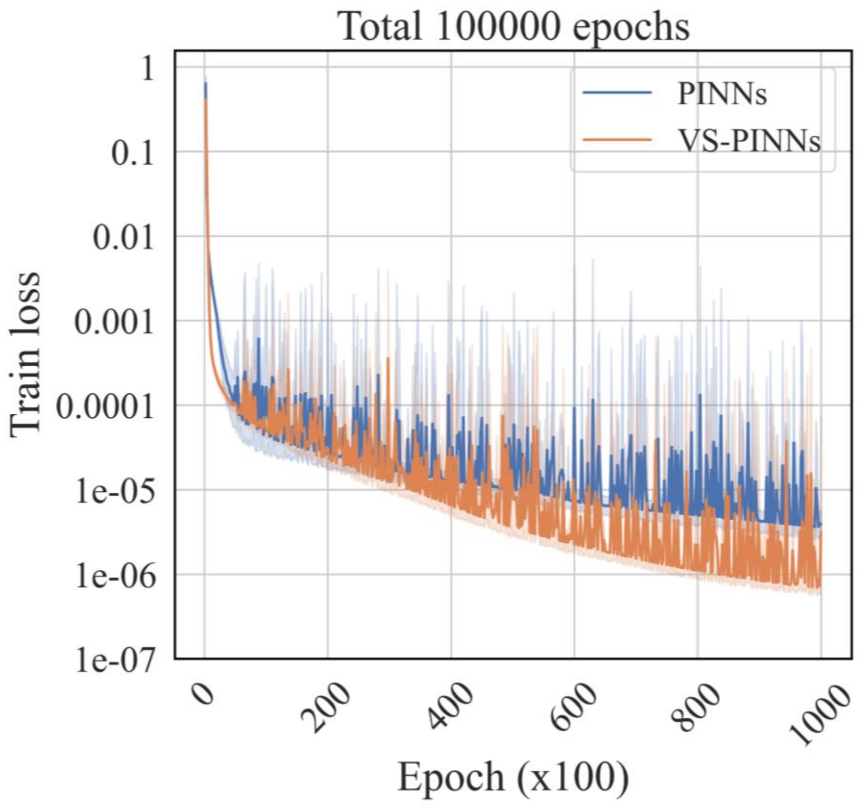


Experiment 3

Convection-Diffusion equation

$$\begin{cases} u_t + u_x - \frac{1}{4}u_{xx} = 0 \\ u(0,x) = \sin(x) \\ u(t,0) = -e^{-\frac{1}{4}t}\sin(t) \\ u(t,\pi) = e^{-\frac{1}{4}t}\sin(\pi - t) \end{cases}$$

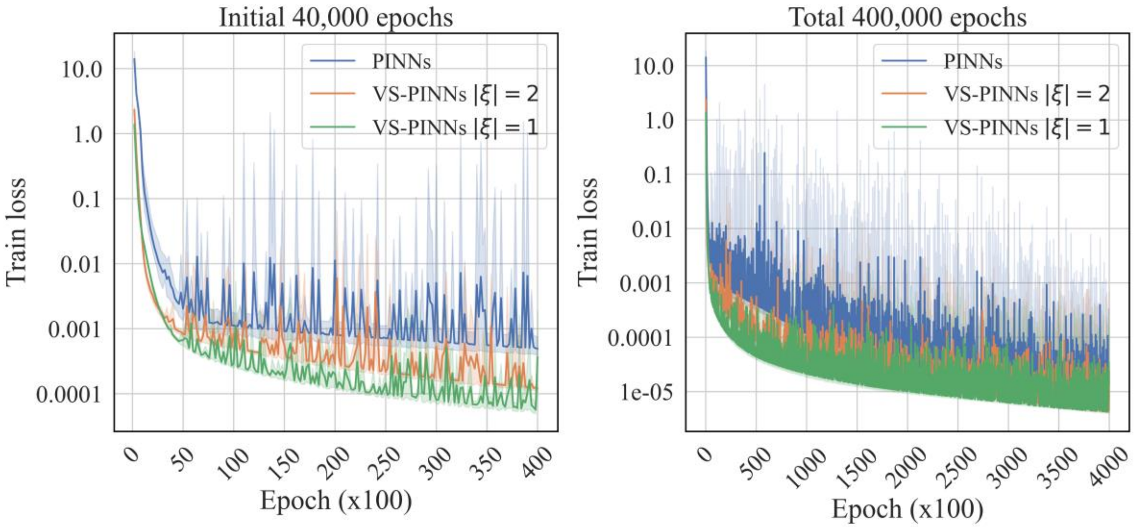




Experiment 4

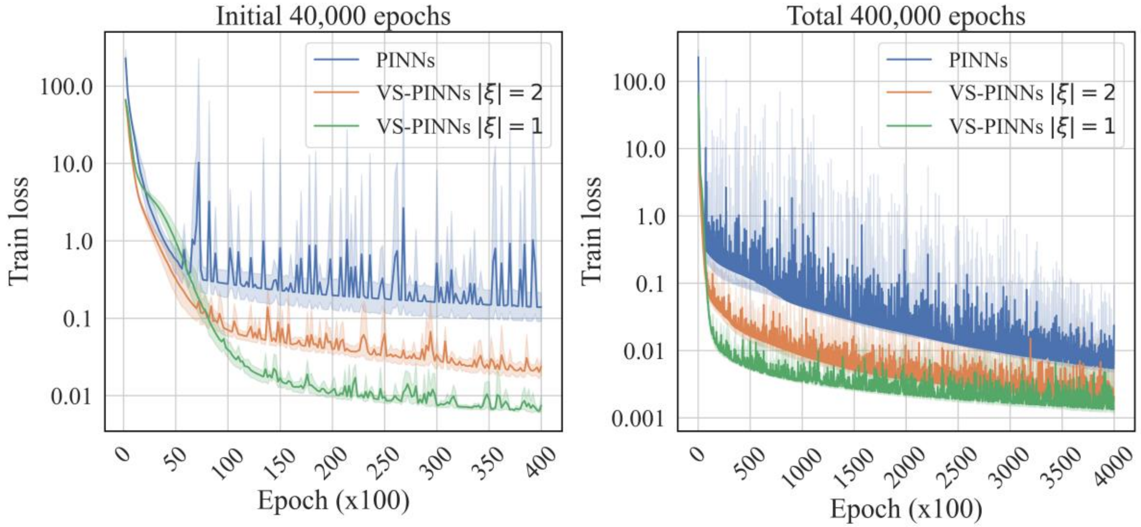
Elastic beam equation

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



Bi-harmonic equation

$$\begin{cases} u_{xxxx} + 2u_{xxyy} + u_{yyyy} = f_2 \\ u(x,0) = u(x,\pi) = u(0,y) = u(\pi,y) = 0 \\ \frac{\partial}{\partial \mathbf{n}} u(x,0) = \frac{\partial}{\partial \mathbf{n}} u(x,\pi) = \frac{\partial}{\partial \mathbf{n}} u(0,y) = \frac{\partial}{\partial \mathbf{n}} u(\pi,y) = 0 \end{cases}$$

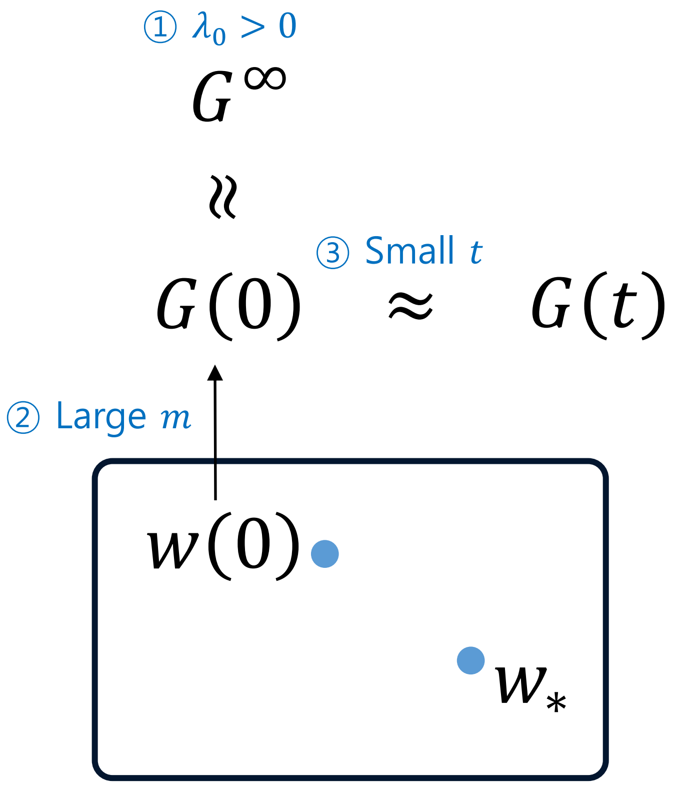


Computational cost

PDE	Method	GPU memory	running time	parameters
Bi-harmonic	PINN	801.052 Mb	0.053 s/epoch	4000
	VS-PINN $\lvert\xi\rvert = 2$	481.094 Mb	0.049 s/epoch	9000
	VS-PINN $\lvert\xi\rvert = 1$	81,466 Mb	0.053 s/epoch	20000
Beam	PINN	323.772 Mb	0.037 s/epoch	4000
	VS-PINN $\lvert\xi\rvert = 2$	240.689 Mb	0.038 s/epoch	9000
	VS-PINN $\lvert\xi\rvert = 1$	80.836 Mb	0.040 s/epoch	17000

- Auxiliary variables need additional models.
- Reducing the number of differentiation in loss is more critical.

Sketch of Proofs



Proposition

$\boldsymbol{G}_{\boldsymbol{v}}^{\infty} = \mathbb{E}_{\boldsymbol{w},\boldsymbol{v}}\left[\boldsymbol{G}_{\boldsymbol{v}}\left(\boldsymbol{w},\boldsymbol{v}\right)\right]$ *is strictly positive definite and independent of m.*

Proposition

For $\delta > 0$ *and some constant* N_1, C_1 *and* R, *if* m *is large enough so that*

$$m \geq \frac{32 N_1 C_1^2 R^{4p}}{\lambda_0^2} \log\left(\frac{2N_1}{\delta}\right),$$

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

$$\left\|\boldsymbol{G}_{\boldsymbol{v}}\left(\boldsymbol{w}(0), \boldsymbol{v}(0)\right) - \boldsymbol{G}_{\boldsymbol{v}}^{\infty}\right\|_2 < \frac{\lambda_0}{4}.$$

Sketch of Proofs

$$\frac{d}{dt}u(t) = G(t)(y - u(t)) \qquad \frac{d}{dt}w_r(t) = -\frac{\partial}{\partial w_r}\|u(t) - y\|_2^2$$

① $\lambda_0 > 0$

$$G^\infty$$

$$\wr\wr$$

③ Small t

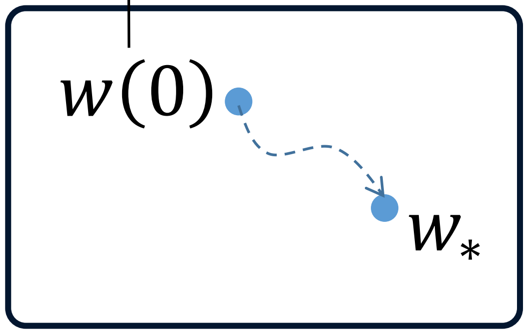
$$G(0) \approx G(t)$$

④ $\lambda_0(t) > \frac{1}{2}\lambda_0$

⑤ Fast decrease $u(t)$

$$\|u(t) - y\|_2 \leq \exp(-\lambda_0 t)\,\|u(0) - y\|$$

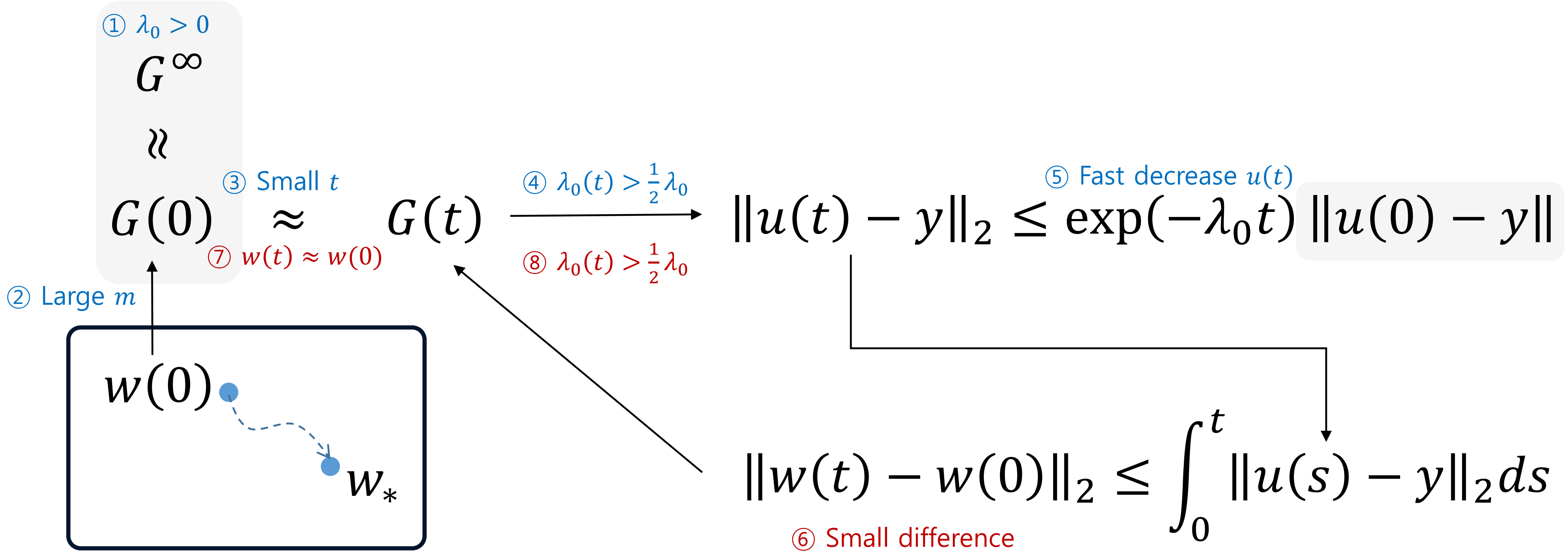
② Large m



$$\|w(t) - w(0)\|_2 \leq \int_0^t \|u(s) - y\|_2 ds$$

Sketch of Proofs

$$\frac{d}{dt}u(t) = G(t)(y - u(t)) \qquad \frac{d}{dt}w_r(t) = -\frac{\partial}{\partial w_r}\|u(t) - y\|_2^2$$



Thank you