AI in the Sciences and Engineering HS 2026: Lecture 4

Siddhartha Mishra

Computational and Applied Mathematics Laboratory (CamLab)
Seminar for Applied Mathematics (SAM), D-MATH (and),
ETH AI Center (and) Swiss National AI Institute (SNAI) ,
ETH Zürich, Switzerland.

What we have learnt so far ?

- AIM: Learn/Solve PDEs using Deep Neural Networks
- Use PINNs for that purpose.

PINNs for the PDE $\mathcal{D}(u) = f$

- For Parameters $\theta \in \Theta$, $u_\theta : \mathbb{D} \mapsto \mathbb{R}^m$ is a DNN, with $u_\theta \in X^*$
- Aim: Find $\theta \in \Theta$ such that $u_\theta \approx u$ (in suitable sense).
- Compute PDE Residual by Automatic Differentiation:

$$\mathcal{R} := \mathcal{R}_\theta(y) = \mathcal{D}(u_\theta(y)) - f(y), \ y \in \mathbb{D} \quad \mathcal{R}_\theta \in Y^*, \quad \forall \theta \in \Theta$$

- PINNs are minimizers of $\|\mathcal{R}_\theta\|_Y^p \sim \int_{\mathbb{D}} |\mathcal{R}_\theta(y)|^p \, dy$
- Replace Integral by Quadrature !
- Let $\mathcal{S} = \{y_i\}_{1 \le i \le N}$ be quadrature points in $\mathbb{D}$, with weights w_i
- PINN for approximating PDE is defined as $u^* = u_{\theta^*}$ such that

$$\theta^* = \arg\min_{\theta \in \Theta} \sum_{i=1}^{N} w_i |\mathcal{R}_\theta(y_i)|^p$$

- Minimize Very high-d Non-Convex loss with ADAM,L-BFGS

Successes I: High-dimensional PDEs

- PINNs for the Heat Equation:

Dimension	Training Error	Total Error
1	2.8×10^{-5}	0.0035%
5	0.0002	0.016%
10	0.0003	0.03%
20	0.006	0.79%
50	0.006	1.5%
100	0.004	2.6%

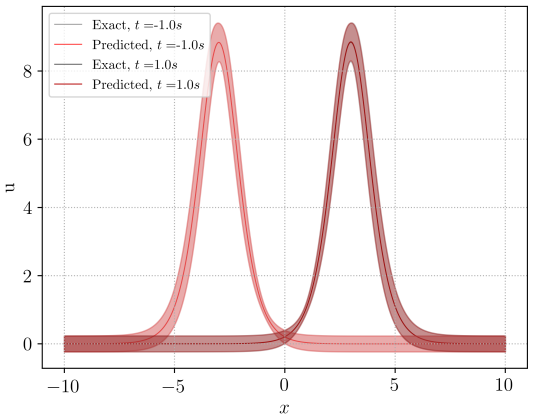
- No Curse of dimensionality !!

Successes II: Parametric PDEs

- Consider the KdV Eqn:

$$u_t + uu_x + u_{xxx} = 0,$$
$$u(0, x, y) = u_0(x, y).$$

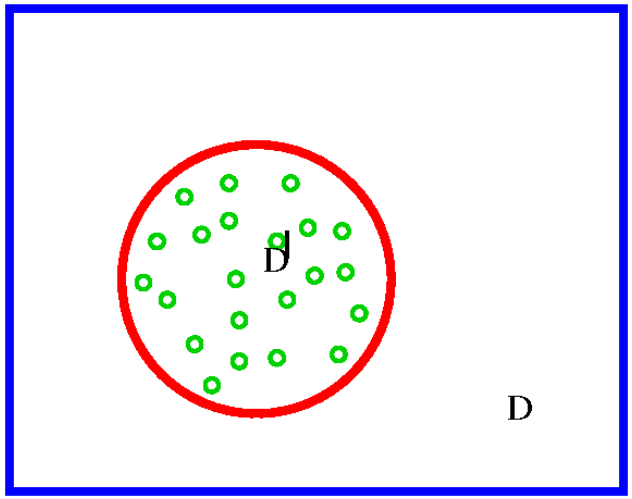
- $y \in Y \subset \mathbb{R}^6$ Parametrizes Initial conditions.
- PINN: $(t, x, y) \mapsto u_\theta(t, x, y)$
- Visualizations of Mean + Variance.

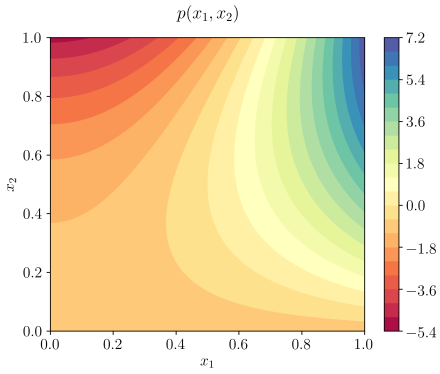


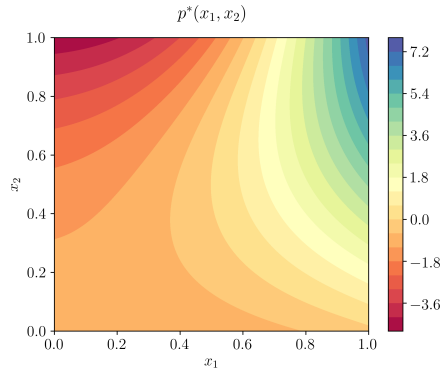
- Error of 0.5%

Success III: Inverse Problems

- Seamless Integration of Data + Physics







Why and When do PINNs work ?

Why do PINNs work for an abstract PDE $\mathcal{D}(u) = f$?

- PDE solution u, DNN u_θ with parameters $\theta \in \Theta$
- AIM is to ensure small Total Error:

$$\mathcal{E}(\theta) := \|u - u_\theta\|_p$$

- PINNs may not have access to samples from Exact Solution u
- On the other hand, PINNs minimize PDE Residual:

$$\mathcal{E}_G(\theta) = \|\mathcal{R}_\theta\|_p = \|\mathcal{D}(u_\theta) - f\|_p$$

- In practice, we only have access to Training Error:

$$\mathcal{E}_T(\theta) = \left(\sum_{i=1}^{N} w_i |\mathcal{R}_\theta(y_i)|^p \right)^{\frac{1}{p}}$$

Key Theoretical Questions

- Can the training error be made as small as possible ?

$$\text{Does } \exists \tilde{\theta} \in \Theta, \quad \mathcal{E}_T(\tilde{\theta}) < \epsilon \text{ ?.}$$

- Does small Training Loss $\Rightarrow$ small PINN Residual ? i.e.,
- Can we derive a bound of the form ?

$$\mathcal{E}_G(\theta) \leq \overline{C}\left(\mathcal{E}_T(\theta), N\right) \sim o\left(N^{-1}\right) \quad \forall \theta \in \Theta$$

- Does small PINN Residual $\Rightarrow$ small Total Error ? i.e.,
- Can we derive a bound of the form:

$$\mathcal{E}(\theta) \leq C\mathcal{E}_G(\theta), \quad \forall \theta \in \Theta$$

On bounds on total error in terms of Residuals

- Sufficient Conditions of SM, Molinaro, 2021:
- Coercivity of the PDE $\mathcal{D}\mathsf{u} = f$: for any $\mathsf{u}, \bar{\mathsf{u}} \in X^*$:

$$\|\mathsf{u} - \bar{\mathsf{u}}\|_p \leq C_{pde}(\bar{\mathsf{u}}, \mathsf{u})\|\mathcal{D}(\bar{\mathsf{u}}) - \mathcal{D}(\mathsf{u})\|_p$$

- Coercivity $\Rightarrow$ Bounds in terms of Residuals as,

$$\begin{aligned}
\mathcal{E}(\theta) &= \|\mathsf{u}_\theta - \mathsf{u}\|_p, \\
&\leq C_{pde}(\mathsf{u}, \mathsf{u}_\theta)\|\mathcal{D}(\mathsf{u}_\theta) - \mathcal{D}(\mathsf{u})\|_p \quad (\text{Coercivity}), \\
&\leq C_{pde}(\mathsf{u}, \mathsf{u}_\theta)\|\mathcal{D}(\mathsf{u}_\theta) - f\|_p \quad \text{as } \mathcal{D}(\mathsf{u}) = f, \\
&\leq C_{pde}(\mathsf{u}, \mathsf{u}_\theta)\mathcal{E}_G(\theta) \quad (\text{Definition of } \mathcal{E}_G)
\end{aligned}$$

On Bounds of Residual in terms of Training Error

- Recall PDE Residual:

$$\mathcal{E}_G(\theta) = \|\mathcal{R}_\theta\|_p = \|\mathcal{D}(\mathsf{u}_\theta) - \mathsf{f}\|_p := \left(\int_D |\mathcal{R}_\theta(y)|^p dy \right)^{\frac{1}{p}}$$

- In practice, we only have access to Training Error:

$$\mathcal{E}_T(\theta) = \left(\sum_{i=1}^{N} w_i |\mathcal{R}_\theta(y_i)|^p \right)^{\frac{1}{p}}$$

- Training Error $\mathcal{E}_T$ is Quadrature Approximation of $\mathcal{E}_G$:

$$\mathcal{E}_G \leq \mathcal{E}_T + C_{quad}(u_{\theta^*})^{\frac{1}{p}} N^{-\frac{\alpha}{p}} \quad \text{quadrature error},$$

A Strategy for PINN Error Bounds SM, Molinaro, 2021

- Use Coercivity of a given PDE to show that

$$\|u - u_\theta\|_p \leq \overline{C}(u, u_\theta)\mathcal{E}_G(\theta), \quad \forall \theta \in \Theta.$$

- Use Quadrature bounds to show that,

$$\mathcal{E}_G \leq \mathcal{E}_T + C_{quad}(u_{\theta^*})^{\frac{1}{p}} N^{-\frac{\alpha}{p}}$$

- Prove explicit growth bounds on the constants $\overline{C}, C_{quad}$ in terms of Neural Network architecture and number of collocation points.

Kolmogorov PDEs

- Linear Parabolic PDEs of form:

$$\partial_t u = \sum_{i=1}^{d} \mu_i(x)\partial_{x_i} u + \frac{1}{2}\sum_{i,j,k=1}^{d} \sigma_{ik}(x)\sigma_{kj}(x)\partial_{x_i x_j} u,$$

$$u|_{\partial D\times(0,T)} = \Psi(x,t), \quad u(x,0) = \varphi(x)$$

- μ, σ are Affine
- Examples:
  - Heat Equation: $\mu = 0$, $\sigma = ID$
  - Black-Scholes Equation for Option Pricing:
  - Interest rate μ, Stock Volatilities β and correlations ρ

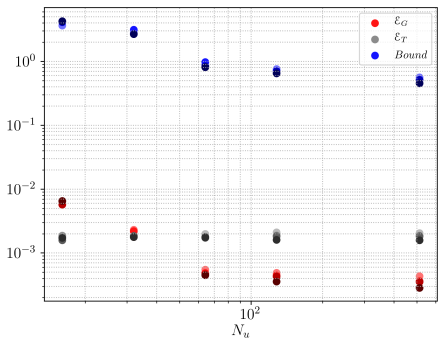
$$u_t = \sum_{i,j=1}^{d} \beta_i\beta_j\rho_{ij}x_i x_j u_{x_i x_j} + \sum_{j=1}^{d} \mu x_j u_{x_j}$$

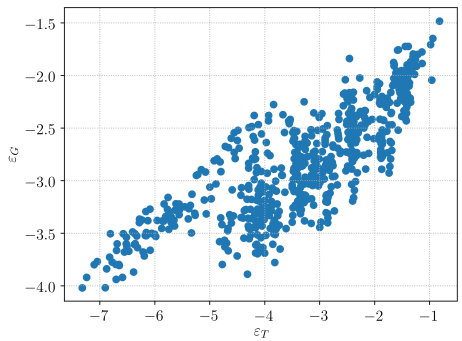
- Note that $d >> 1$ (Very high-dimensional)

Error Bounds: De Ryck, SM, 2021.

- $\exists$ Tanh PINN $\hat{u}$ of size $\mathcal{O}(\epsilon^{-\alpha(d)})$: $\mathcal{E}_{G,T}(\hat{\theta}) \sim \epsilon$,
- Uses Dynkin's formula to overcome curse of dimensionality.
- Stability of PDE: $\|u - u_\theta\|_2 \leq C\left(\|\mathcal{R}_{int,\theta}\| + \|\mathcal{R}_{sb,\theta}\|^{\frac{1}{2}}\right)$
- Use Hoeffding's inequality + Lipschitz bounds on u_θ:

$$\mathcal{E}_G^2(\theta) \sim \mathcal{O}\left(\mathcal{E}_T^2(\theta) + \frac{C\left(M, \log(\|W\|)\right)\log(\sqrt{N})}{\sqrt{N}}\right)$$





Numerical Results: (SM, Molinaro, Tanios, 2021)

- Heat Equation:

Dimension	Training Error	Total error
20	0.006	0.79%
50	0.006	1.5%
100	0.004	2.6%

- Black-Scholes type PDE with Uncorrelated Noise:

Dimension	Training Error	Total error
20	0.0016	1.0%
50	0.0031	1.5%
100	0.0031	1.8%

- Heston option-pricing PDE

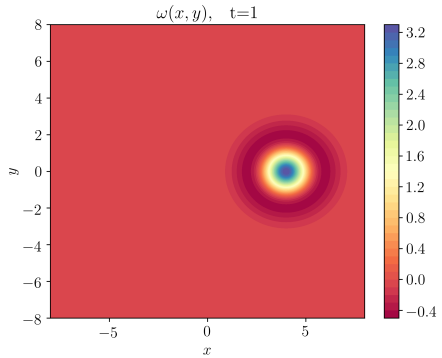
Dimension	Training Error	Total error
20	0.0064	1.0%
50	0.0037	1.3%
100	0.0032	1.4%

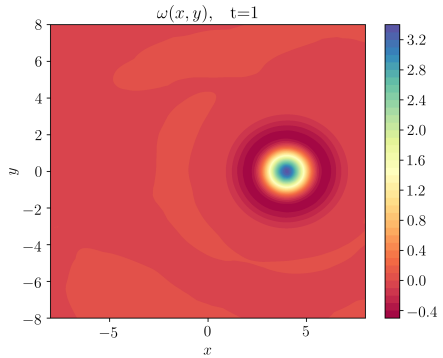
Navier-Stokes Eqns: $\mathrm{u}_t + (\mathrm{u}\cdot\nabla)\mathrm{u} + \nabla p = \nu\Delta u,\ \mathrm{div}\ \mathrm{u} = 0$

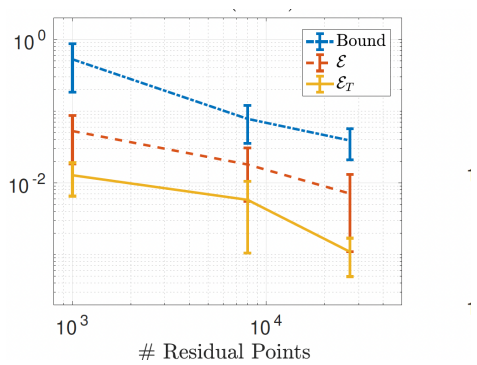
- Theory in DeRyck, Jagtap, SM, 2022.
- Smooth $u \in H^k$: PINN with size$(\hat{u}) \sim \mathcal{O}\left(M^{d+1}\right)$: $\mathcal{E}_G(\hat{\theta}) \leq \mathcal{O}\left(M^{1-k}\log(M)\right)$
- Use PDE theory to prove for $C = C\left(\|\mathrm{curl}\ u\|_{L^\infty}\right)$

$$\|u - u_\theta\|_2 \leq C\left(\|\mathcal{R}_{int,\theta}\| + \|\mathcal{R}_{tb,\theta}\| + \|\mathcal{R}_{sb,\theta}\|^{\frac{1}{2}} + \|\mathcal{R}_{div,\theta}\|^{\frac{1}{2}}\right)$$

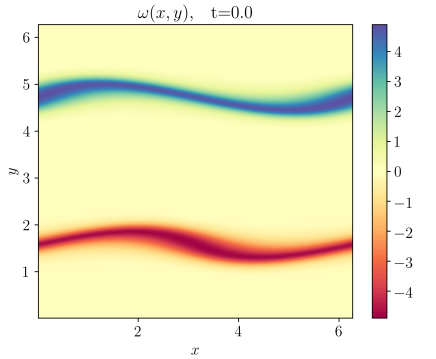
- Use Quadrature bounds: $\mathcal{E}_G^2(\theta) \sim \mathcal{O}\left(\mathcal{E}_T^2(\theta) + N^{-\alpha}\right)$

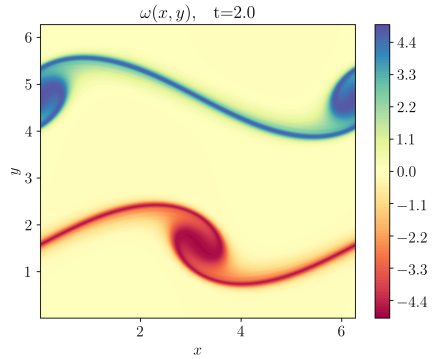


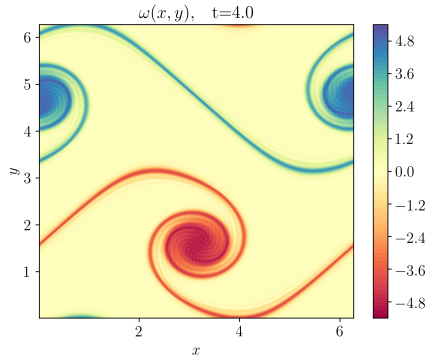


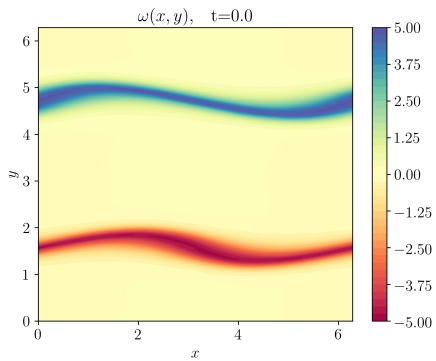


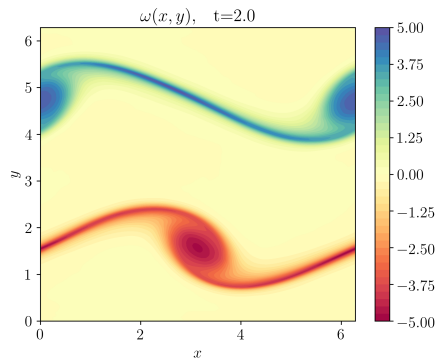
Results for 2-D Double Shear Layer

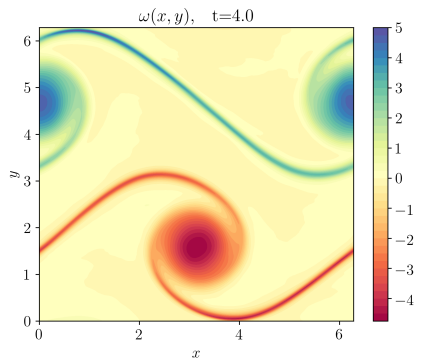






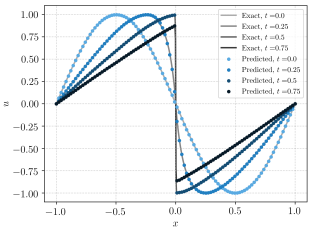




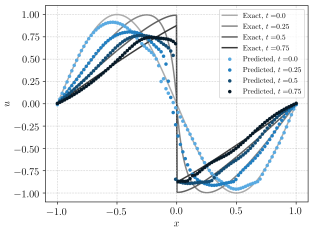


Viscous Burgers': $u_t + \text{div}\ f(u) = \nu \Delta u$

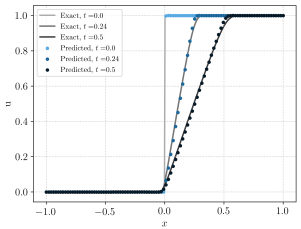
- Error $\mathcal{E} \leq Ce^{CT}\left(\mathcal{E}_T + C_q N^{-\alpha}\right)$, $C = C\left(\|\nabla u^\nu\|_{L^\infty}\right)$
- $\|\nabla u^\nu\|_{L^\infty} \sim \frac{1}{\sqrt{\nu}} \Rightarrow$ Error can blow up near shocks !!



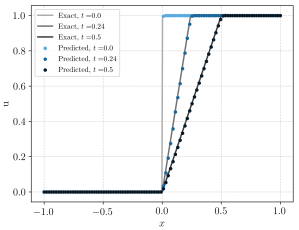
$\nu = 10^{-3}$, Sh



$\nu = 0$, Sh



$\nu = 10^{-3}$, RF



$\nu = 0$, RF

ν	$\mathcal{E}$ (Shock)	$\mathcal{E}$ (Rarefaction)
10^{-3}	1.0%	2.2%
10^{-4}	11.2%	1.6%
0	23.1%	1.2%

- Alternatives: wPINNs of De Ryck, Molinaro, SM, 2023.

Summary (so far)

- For generic PDE: $\mathcal{D}(\mathrm{u}) = f$
- Rigorous Error estimate for PINNs:

$$\|\mathrm{u} - \mathrm{u}_\theta\| \sim C_{\mathrm{pde}}(\mathrm{u}, \mathrm{u}_\theta)\left[\mathcal{E}_T(\theta) + C_{\mathrm{quad}}(\mathrm{u}_\theta)N^{-\alpha}\right]$$

- Training Error is a blackbox

On the smallness of Training Error

- For sufficiently smooth u solving $\mathcal{D}(\mathrm{u}) = \mathrm{f}$ observe that

$$\mathcal{E}_G(\theta) = \|\mathcal{D}(\mathrm{u}_\theta) - \mathrm{f}\|_p = \|\mathcal{D}(\mathrm{u}_\theta) - \mathcal{D}(\mathrm{u})\|_p \leq C(\mathrm{u}, \mathrm{u}_\theta) \|\mathrm{u} - \mathrm{u}_\theta\|_{W^{s,p}}$$

- Here s depends on the number of derivatives in $\mathcal{D}$.
- DNN approximation results in Sobolev spaces:
- For example in (DeRyck,Lanthaler,SM, 2021) :

$$\exists \hat{\theta} \in \Theta, \quad \|\mathrm{u} - \mathrm{u}_{\hat{\theta}}\|_{W^{s,p}} < \epsilon$$

- smoothness of u $\Rightarrow$ small PINN Residuals.
- Use Quadrature bounds to show that,

$$\mathcal{E}_T \leq \mathcal{E}_G + C_{quad}(u_{\theta^*})^{\frac{1}{p}} N^{-\frac{\alpha}{p}} \sim O(\epsilon), \text{ for } N >> 1,$$

On the smallness of Training Error

- For sufficiently smooth u solving $\mathcal{D}(\mathrm{u}) = \mathrm{f}$ observe that

$$\mathcal{E}_G(\theta) = \|\mathcal{D}(\mathrm{u}_\theta) - \mathrm{f}\|_p = \|\mathcal{D}(\mathrm{u}_\theta) - \mathcal{D}(\mathrm{u})\|_p \leq C(\mathrm{u}, \mathrm{u}_\theta) \|\mathrm{u} - \mathrm{u}_\theta\|_{W^{s,p}}$$

- Here s depends on the number of derivatives in $\mathcal{D}$.
- DNN approximation results in Sobolev spaces:
- For example in (DeRyck,Lanthaler,SM, 2021) :

$$\exists \hat{\theta} \in \Theta, \quad \|\mathrm{u} - \mathrm{u}_{\hat{\theta}}\|_{W^{s,p}} < \epsilon$$

- smoothness of u $\Rightarrow$ small PINN Residuals.
- Use Quadrature bounds to show that,

$$\mathcal{E}_T \leq \mathcal{E}_G + C_{quad}(u_{\theta^*})^{\frac{1}{p}} N^{-\frac{\alpha}{p}} \sim O(\epsilon), \text{ for } N >> 1,$$

- Can training process reach the Global Minimum ?

Theoretical Framework of De Ryck et al 2023

- Gradient Descent with Physics-Informed Loss:

$$\theta_{k+1} = \theta_k - \eta \nabla_\theta L, \quad L = \frac{1}{2} \int_D |\mathcal{D}(\mathrm{u}(x,\theta) - f(x)|^2 dx.$$

- Taylor Expansion:

$$\mathrm{u}(x,\theta_k) = \mathrm{u}(x,\theta_0) + \nabla_\theta \mathrm{u}(x,\theta_0)(\theta_k - \theta_0) + \langle H_k \theta_k - \theta_0, \theta_k - \theta_0 \rangle$$

- Rewritten GD: $\theta_{k+1} = (I - \eta \mathcal{A})\theta_k + \eta(\mathcal{A}\theta_0 + \mathcal{C}) + \eta \epsilon_k$
- Gram Matrix: $\mathcal{A}_{i,,j} = \langle \mathcal{D}\varphi_i, \mathcal{D}\varphi_j \rangle_{L^2}$, $\varphi_i = \partial_{\theta_i} \mathrm{u}(x,\theta_0)$
- Bias vector: $\mathcal{C}_i = \langle \mathcal{D}\mathrm{u}(\theta_0) - f, \mathcal{D}\varphi_i \rangle$

Dynamics of simplified GD

- if $\epsilon_k \sim \mathcal{O}(\epsilon)$, then *GD* can be approximated by simpGD:

$$\theta_{k+1} = (I - \eta\mathcal{A})\theta_k + \eta(\mathcal{A}\theta_0 + \mathcal{C})$$

- Small error terms correspond to the NTK regime for $\mathrm{u}_\theta, \mathcal{D}\mathrm{u}_\theta$:

$$TKf_\theta(x,y) = \nabla_\theta f_\theta(x)^\top \nabla_\theta f(y).$$

- For simpGD, easy to show that

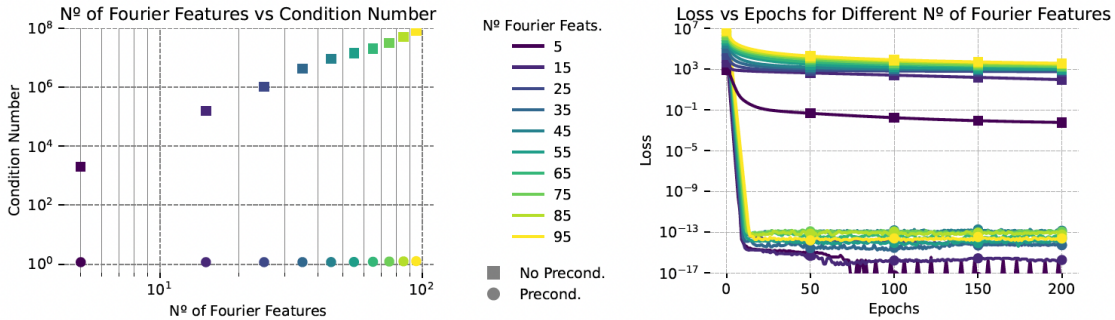
$$\|\theta_k - \theta^*\|_2 \leq \left(1 - \frac{c}{\kappa(\mathcal{A})}\right)^k \|\theta_0 - \theta^*\|_2, \quad N(\delta) \sim \mathcal{O}(\kappa(\mathcal{A}) \log(1/\delta))$$

- Key role played by Condition Number: $\kappa(\mathcal{A}) = \frac{\lambda_{\max}(\mathcal{A})}{\lambda_{\min}(\mathcal{A})}$

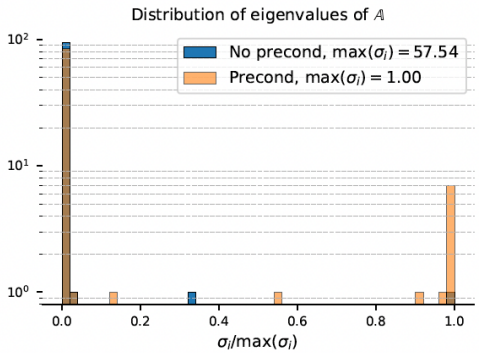
More on Simp GD

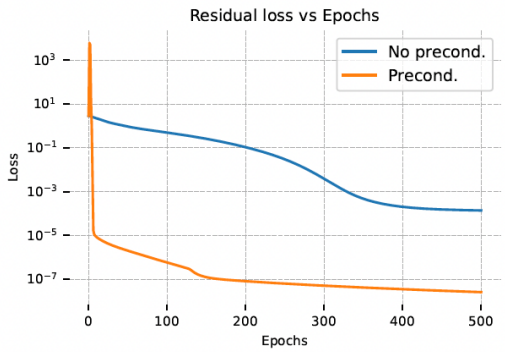
- Introduce $\mathcal{A} = \mathcal{D}^*\mathcal{D}$, the Hermitian-Square of $\mathcal{D}$.
- Under suitable assumptions, $\kappa(\mathcal{A}) = \kappa\left(\mathcal{A} \odot TT^*\right)$,
- $T : v \mapsto \sum_k v_k \varphi_k$ connects the vector and function spaces.
- Ex: if $\mathcal{D} = -\Delta$, then $\mathcal{A} = \Delta^2$
- in general $\kappa(\mathcal{A})$ can be very high.
- Key difference in Supervised Learning and Physics-Informed learning
- Need to precondition $\mathcal{D}^*\mathcal{D}$.
- Most techniques to accelerate PINNs training can be viewed as Preconditioning

1-D Possion: $-u'' = -k^2 \sin(kx)$

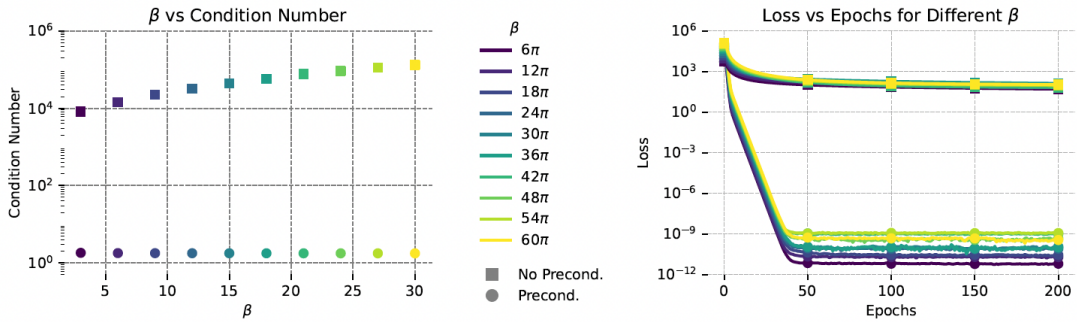


1-D Possion: $-u'' = -k^2 \sin(kx)$

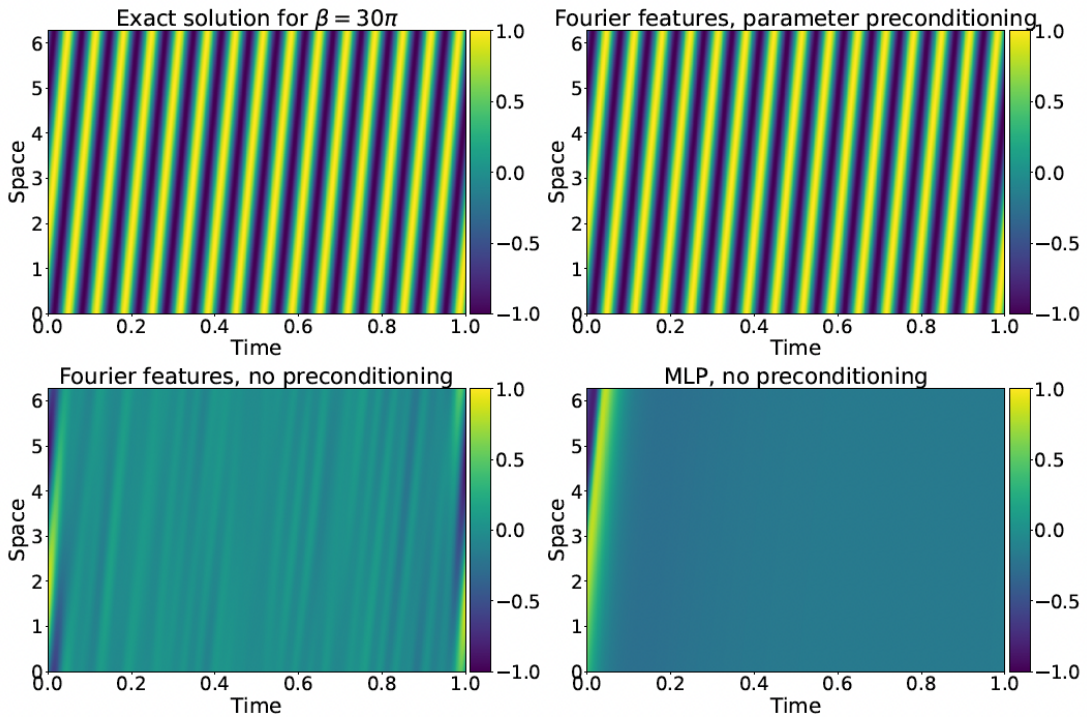




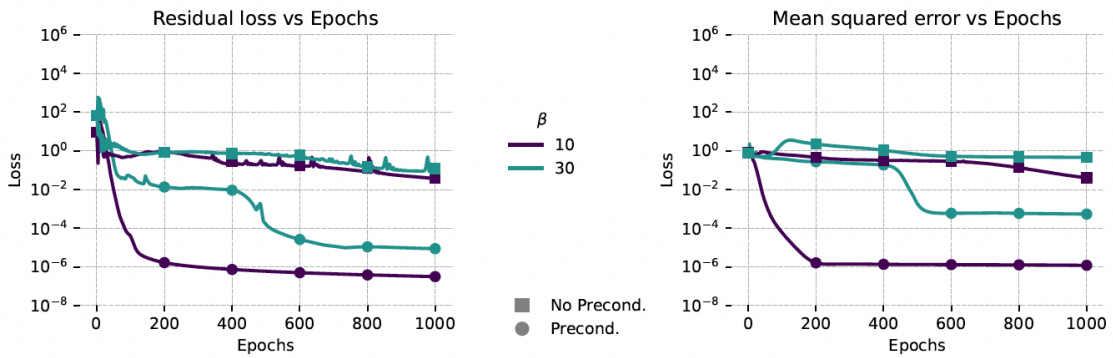
1-D Advection: $u_t + \beta u_x = 0$



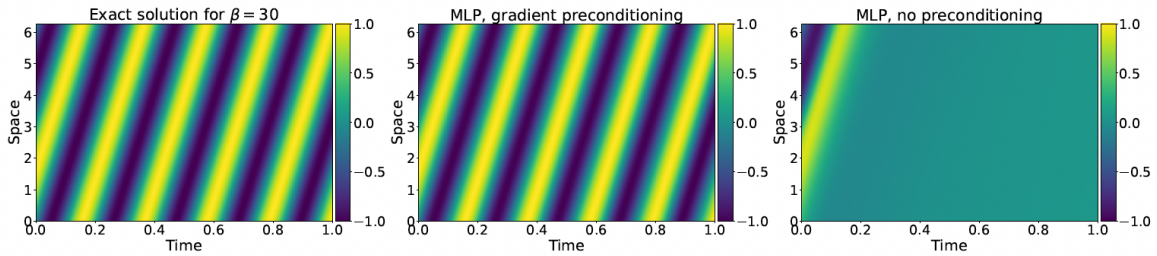
1-D Advection: $u_t + \beta u_x = 0$



1-D Advection: $u_t + \beta u_x = 0$



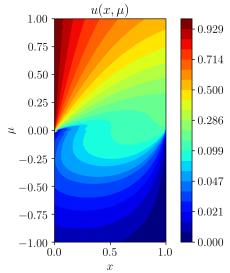
1-D Advection: $u_t + \beta u_x = 0$



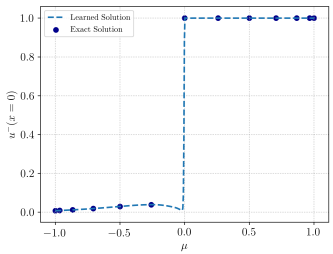
Preconditioning Techniques

- Adjusting balance between Data vs. Physics
- Hard Boundary Conditions (Lagaris et. al)
- Casual Learning (Wang, et. al)
- Second-order Optimizers (Zeinhofer, et. al)
- Multi-stage Neural Networks (Lai et. al)

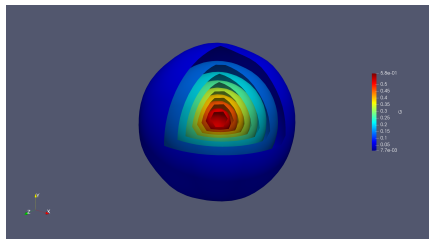
Radiative Transfer



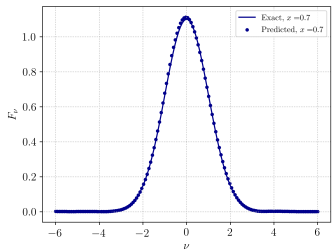
2-D, Intensity



2-D, Boundary



6-D, Inc. Radiation

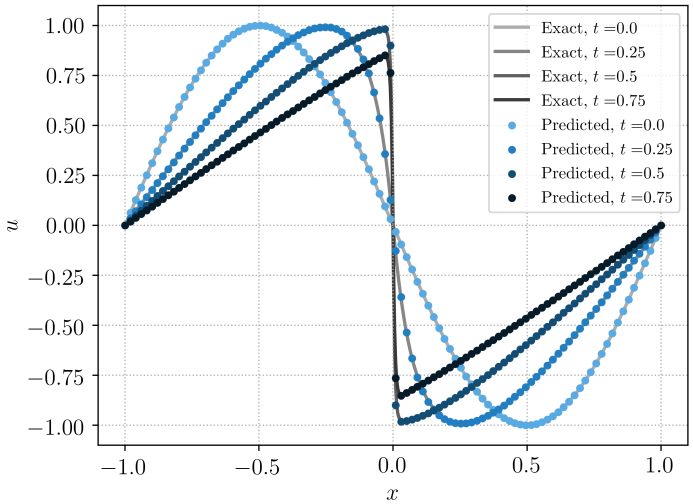


6-D, Radial flux

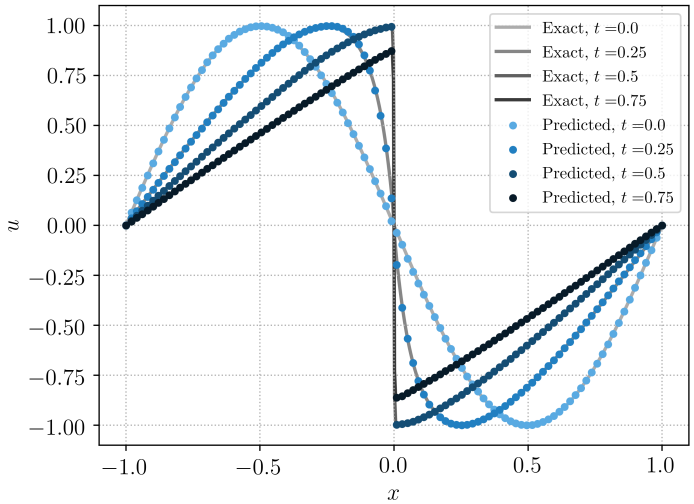
Dimension	Network Size	Error	Training Time
2	24×8	0.3%	57 min
6	20×8	2.1%	66 min

Results for 1-D Burgers'

- Sobol points, $N_{int} = 8192$, $N_{tb} = N_{sb} = 256$, Depth 8, Width 20.



$\nu = 10^{-2}$

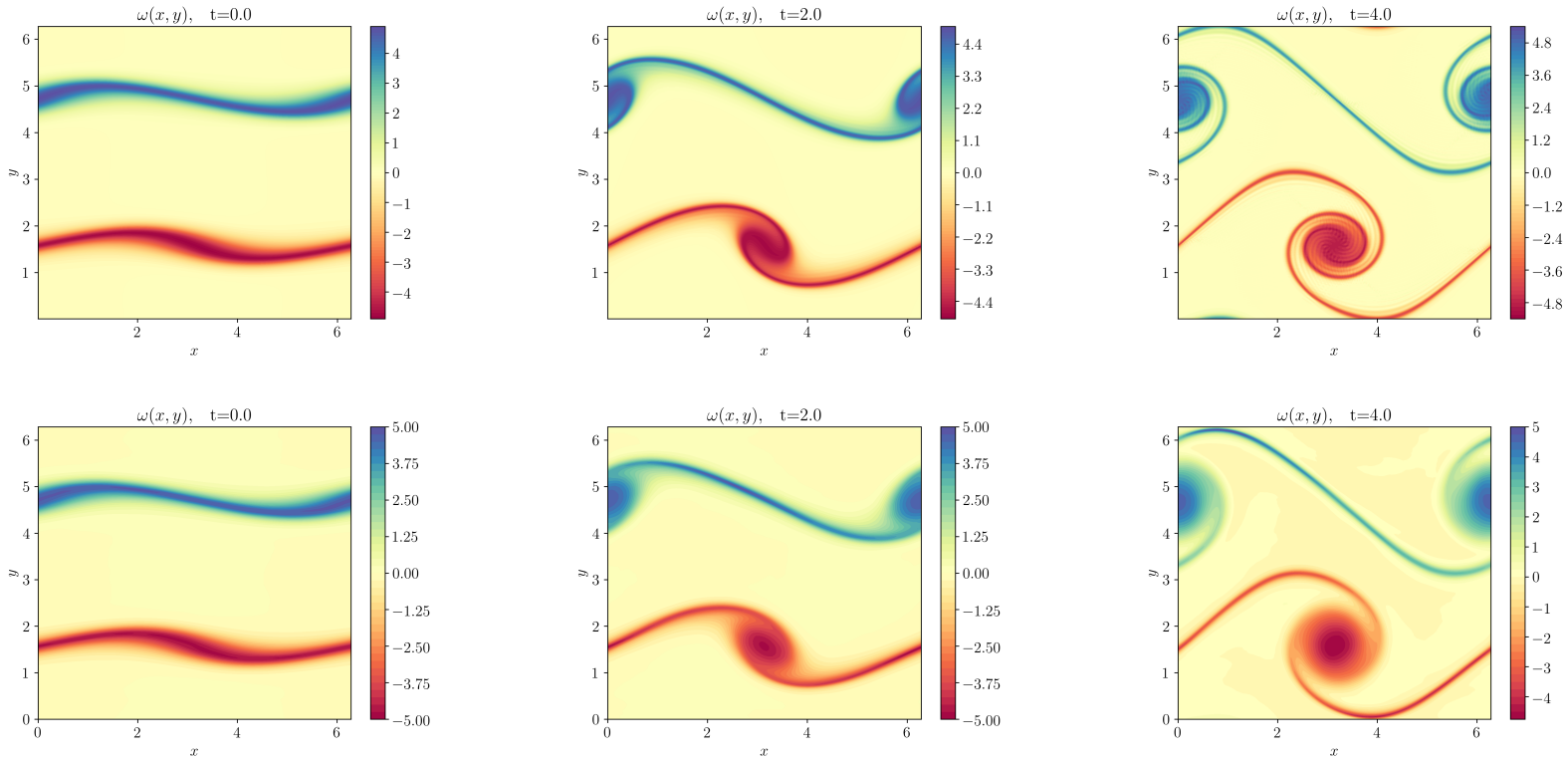


$\nu = 10^{-3}$

Viscosity	Training Error	Total error
10^{-2}	0.0005	1.0%
10^{-3}	0.0008	1.2%

- Finite Difference (0.1 secs) vs. PINNs (5 min)

Results for 2-D Navier-Stokes



- Spectral Method (1 secs) vs. PINNs (30-60 min)

Summary of PINNs.

- PINNs are alternatives to PDE Solvers.
- Work well for problems with "easy" solutions.
- Don't work yet for complex problems.
- Whats the alternative ?