Brief review on failure mode of PINNs

HAN Jiayue

Nov 2022

From a reviewer...

Why we need to study the fail mode of PINNs?

Weaknesses:
A well known failure mode of PINNs is in stiff PDE domains. Authors have not explicitly demonstrated or characterized their method in this important context (i.e., in stiff PDE contexts) as this is one of the main application areas where such re-weighting schemes can greatly help PINN models avoid pathologies and consequent convergence to trivial and inconsistent solutions.

Figure: Reviewer's opinion

Introduction

We introduce 5 papers that deal with the failure mode of PINN.

- Characterizing possible failure modes in PINNs. (Warm start+seq2seq)
- Lagrangian PINNs: A causality-conforming solution to failure modes of physics-informed neural networks. (Lagrangian PINNs)
- Critical Investigation of Failure Modes in PINNs. (PECANN)
- Investigating and Mitigating Failure Modes in PINNs.
- Mitigating Propagation Failures in PINNs using evolutionary sampling. (Evo)

Characterizing possible failure modes in PINNs

- Nov 2021 (early work).
- Discussed the failure mode in details.
- Proposed 2 methods: Warm start+seq2seq.

Failure mode

Consider a one-dimensional convection problem

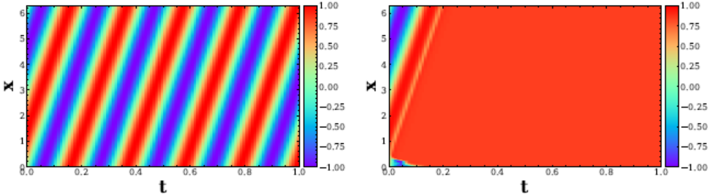
$$\frac{\partial u}{\partial t}+\beta\frac{\partial u}{\partial x}=0,\quad x\in\Omega, t\in[0,T]$$
$$u(x,0)=h(x),\quad x\in\Omega.$$

Here, β is the convection coefficient and $h(x)$ is the initial condition. For constant β and periodic boundary conditions, this problem has a simple analytical solution:

$$u_{\text{analytical}}(x,t)=F^{-1}\left(F(h(x))e^{-i\beta kt}\right)$$

Failure mode

The PINN fails when β becomes larger, reaching a relative error of almost 100% for $\beta > 10$.



(b) *Exact solution for* $\beta = 30$ **(c)** *PINN solution for* $\beta = 30$

Figure: Failure mode

Failure mode

Samilar mode can also be found in the reaction-diffusion system:

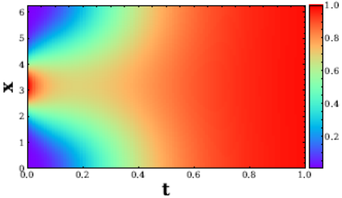
$$\frac{\partial u}{\partial t} - \nu \frac{\partial^2 u}{\partial x^2} - \rho u(1-u) = 0, \quad x \in \Omega, t \in (0, T],$$
$$u(x,0) = h(x), \quad x \in \Omega.$$

The diffusion equation has the following analytical solution:

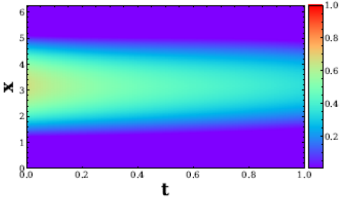
$$u_{\text{analytical}}(x,t) = F^{-1}\left(F\left(u\left(x, t = t^n\right)\right) e^{-\nu k^2 t}\right),$$

where $u\left(x, t = t^n\right)$ is the solution at the n^{th} time step.

Failure mode



(a) *Exact solution for* $\rho = 5, \nu = 5$



(b) *PINN solution for* $\rho = 5, \nu = 5$

Figure: Failure mode

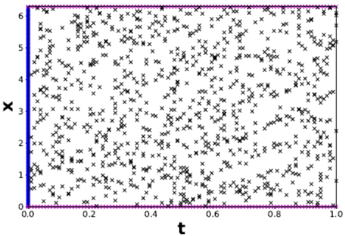
Curriculum regularization

We devise a "curriculum regularization" method to warm start the NN training by finding a good initialization for the weights.
Instead of training the PINN to learn the solution right away for cases with higher β/ρ, we start by training the PINN on lower β/ρ (easier for the PINN to learn) and then gradually move to training the PINN on higher β/ρ, respectively.
(ρ is the coefficient for $u(1-u)$ term)

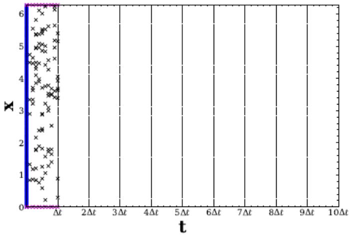
Curriculum regularization

		Regular PINN	Curriculum training
1D convection: $\beta = 20$	Relative error	7.50×10^{-1}	$\mathbf{9.84 \times 10^{-3}}$
	Absolute error	4.32×10^{-1}	$\mathbf{5.42 \times 10^{-3}}$
1D convection: $\beta = 30$	Relative error	8.97×10^{-1}	$\mathbf{2.02 \times 10^{-2}}$
	Absolute error	5.42×10^{-1}	$\mathbf{1.10 \times 10^{-2}}$
1D convection: $\beta = 40$	Relative error	9.61×10^{-1}	$\mathbf{5.33 \times 10^{-2}}$
	Absolute error	5.82×10^{-1}	$\mathbf{2.69 \times 10^{-2}}$

Seq2Seq



(a) *Regular PINN training*



(b) *Sequence-to-sequence learning (model trained every Δt)*

Seq2Seq

		Entire state space	$\Delta t = 0.05$	$\Delta t = 0.1$
$\nu = 2, \rho = 5$	Relative error	5.07×10^{-1}	2.04×10^{-2}	$\mathbf{1.18 \times 10^{-2}}$
	Absolute error	2.70×10^{-1}	1.06×10^{-2}	$\mathbf{6.41 \times 10^{-3}}$
$\nu = 3, \rho = 5$	Relative error	7.98×10^{-1}	1.92×10^{-2}	$\mathbf{1.56 \times 10^{-2}}$
	Absolute error	4.79×10^{-1}	1.01×10^{-2}	$\mathbf{8.17 \times 10^{-3}}$
$\nu = 4, \rho = 5$	Relative error	8.84×10^{-1}	2.37×10^{-2}	$\mathbf{1.59 \times 10^{-2}}$
	Absolute error	5.74×10^{-1}	1.15×10^{-2}	$\mathbf{8.01 \times 10^{-3}}$
$\nu = 5, \rho = 5$	Relative error	9.35×10^{-1}	$\mathbf{2.36 \times 10^{-2}}$	2.39×10^{-2}
	Absolute error	6.46×10^{-1}	$\mathbf{1.09 \times 10^{-2}}$	1.15×10^{-2}
$\nu = 6, \rho = 5$	Relative error	9.60×10^{-1}	2.81×10^{-2}	$\mathbf{2.69 \times 10^{-2}}$
	Absolute error	6.84×10^{-1}	$\mathbf{1.17 \times 10^{-2}}$	1.28×10^{-2}

Lagrangian PINNs

- May 2022.
- New network structure.
- Lagrangian used.

Lagrangian PINNs

Consider the following scalar, one-dimensional convection-diffusion equation

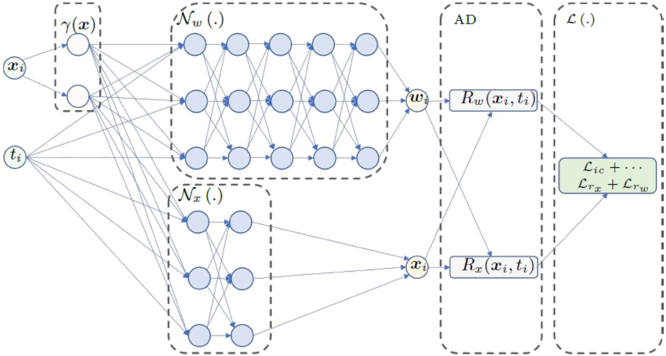
$$R := \frac{\partial w(x,t)}{\partial t} + f_1(x,t,w)\frac{\partial w(x,t)}{\partial x} - f_2(x,t,w)\frac{\partial^2 w(x,t)}{\partial x^2} = 0$$

in the domain $(x,t) \in [x_a, x_b] \times [0, T]$, with initial conditions $w(x,0) = w_0(x)$, and appropriate boundary conditions at x_a and x_b.
Reformulated in the Lagrangian frame gives

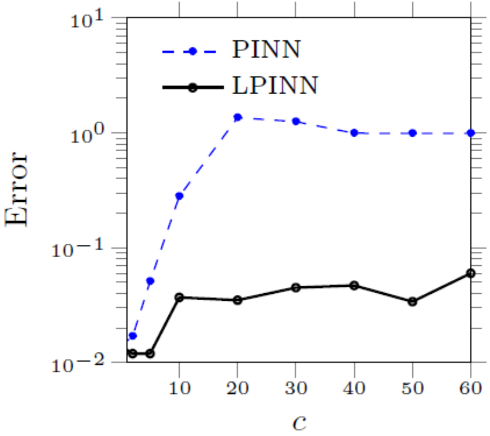
$$\begin{aligned} R_x &:= \frac{dx}{dt} - f_1(x,t,w) = 0 \\ R_w &:= \frac{\partial w}{\partial t} - f_2(x,t,w)\frac{\partial^2 w}{\partial x^2} = 0 \end{aligned}$$

where x is the characteristic curves and w is the state variable on the characteristic curves.

Lagrangian PINNs



Lagrangian PINNs



PECANN

- Jun 2022.
- Lagrangian used.

PECANN

Recall PINN:

$$\mathcal{L}_{\Omega} = \frac{1}{N_{\Omega}} \sum_{i=1}^{N_{\Omega}} \left[\mathcal{D} \left(x^{(i)}, u_{\theta} \left(x^{(i)} \right) \right) \right]^2,$$

$$\mathcal{L}_{\partial\Omega} = \frac{1}{N_{\partial\Omega}} \sum_{j=1}^{N_{\theta\Omega}} \left[g \left(x^{(j)} \right) - u_{\theta} \left(x^{(j)} \right) \right]^2,$$

$$\mathcal{L} = \mathcal{L}_{\Omega} + \mathcal{L}_{\partial\Omega}$$

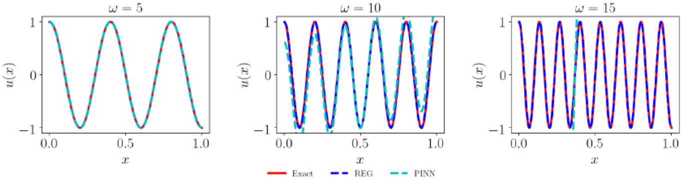
PECANN

Now state the PECANN as follows:

$$
\begin{aligned}
\mathcal{L}_\Omega &= \sum_{i=1}^{N_\Omega} \left[\mathcal{D}\left(x^{(i)}, u_\theta \left(x^{(i)} \right) \right) \right]^2 \\
\mathcal{L}_\mu &= \mathcal{L}_\Omega + \lambda^T \mathcal{C} + \frac{\mu}{2} \|\mathcal{C}\|^2, \\
\lambda &= \lambda + \mu \mathcal{C},
\end{aligned}
$$

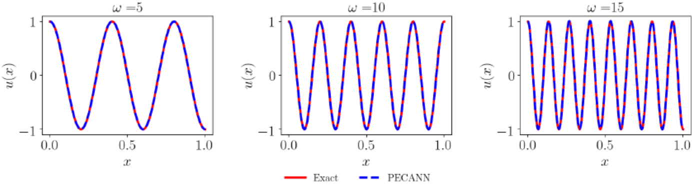
where μ is a positive penalty parameter with a maximum safeguarding value of $\mu_\infty, \lambda^T = \left\{ \lambda^{(1)}, \cdots, \lambda^{(N_{\theta\Omega})} \right\}$ is a vector of Lagrange multipliers, $\mathcal{C}^T = \left\{ \left[g\left(\boldsymbol{x}^{(1)} \right) - u_\theta \left(\boldsymbol{x}^{(1)} \right) \right]^2, \cdots, \left[g\left(\boldsymbol{x}^{(N_{\theta\Omega})} \right) - u_\theta \left(\boldsymbol{x}^{(N_{\theta\Omega})} \right) \right]^2 \right\}$ is a vector of evaluated boundary constraints and T is transposition.

PECANN



(a) (b) (c)

PECANN



(a) (b) (c)

Investigating and Mitigating Failure Modes in PINNs

- Sep 2022.
- Lagrangian used.

Investigating and Mitigating Failure Modes in PINNs

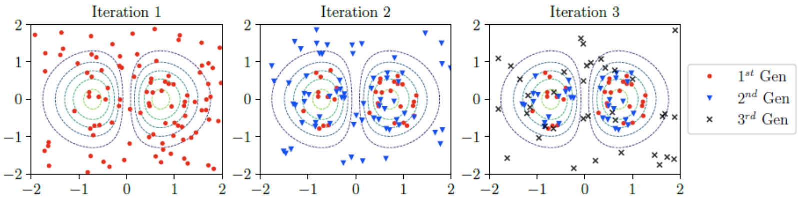
A little bit long, lets's see the paper.

Evo

Related to our method.

- Oct 2022.
- Sampling.

Evo



Evo

Algorithm 1 Proposed Evolutionary Sampling Algorithm For PINN

Sample the initial population $\mathcal{P}_0$ of N_r collocations point $\mathcal{P}_0 \leftarrow \{\mathbf{x_r}\}_{i=1}^{N_r}$ from a uniform distribution $\mathbf{x_r}^i \sim \mathcal{U}(\Omega)$, where Ω is the input domain ($\Omega = [0, T] \times \mathcal{X}$).
for i = 0 to max_iterations - 1 **do**
    Compute the fitness of collocation points $\mathbf{x_r} \in \mathcal{P}_i$ as $\mathcal{F}(\mathbf{x_r}) = |\mathcal{R}(\mathbf{x_r})|$.
    Compute the threshold $\tau_i = \frac{1}{N_r} \sum_{j=1}^{N_r} \mathcal{F}(\mathbf{x_r}^j)$
    Select the retained population $\mathcal{P}_i^r$ such that $\mathcal{P}_i^r \leftarrow \{\mathbf{x_r}^j : \mathcal{F}(\mathbf{x_r}^j) > \tau_{it}\}$
    Generate the re-sampled population $\mathcal{P}_i^s \leftarrow \{\mathbf{x_r}^j : \mathbf{x_r}^j \sim \mathcal{U}(\Omega)\}$
    Merge the two populations $\mathcal{P}_{i+1} \leftarrow \mathcal{P}_i^r \cup \mathcal{P}_i^s$
end for

Evo

Epochs.	Convection ($\beta = 30$) 100k	300k	Convection ($\beta = 50$) 150k	300k	Allen Cahn 200k
PINN (fixed)	$107.5 \pm 10.9\%$	$107.5 \pm 10.7\%$	$108.5 \pm 6.38\%$	$108.7 \pm 6.59\%$	$69.4 \pm 4.02\%$
PINN (dynamic)	$2.81 \pm 1.45\%$	$1.35 \pm 0.59\%$	$24.2 \pm 23.2\%$	$56.9 \pm 9.08\%$	$0.77 \pm 0.06\%$
Curr Reg	$63.2 \pm 9.89\%$	$2.65 \pm 1.44\%$	$48.9 \pm 7.44\%$	$31.5 \pm 16.6\%$	-
CPINN (fixed)	$138.8 \pm 11.0\%$	$138.8 \pm 11.0\%$	$106.5 \pm 10.5\%$	$106.5 \pm 10.5\%$	$48.7 \pm 19.6\%$
CPINN (dynamic)	$52.2 \pm 43.6\%$	$23.8 \pm 45.1\%$	$79.0 \pm 5.11\%$	$73.2 \pm 8.36\%$	$1.5 \pm 0.75\%$
RAR-G	$10.5 \pm 5.67\%$	$2.66 \pm 1.41\%$	$65.7 \pm 17.0\%$	$43.1 \pm 28.9\%$	$25.1 \pm 23.2\%$
RAD	$3.35 \pm 2.02\%$	$1.85 \pm 1.90\%$	$66.0 \pm 1.55\%$	$64.1 \pm 1.98\%$	$0.78 \pm 0.05\%$
RAR-D	$67.1 \pm 4.28\%$	$32.0 \pm 25.8\%$	$82.9 \pm 5.99\%$	$75.3 \pm 9.58\%$	$51.6 \pm 0.41\%$
L^∞	$66.6 \pm 2.35\%$	$41.2 \pm 27.9\%$	$76.6 \pm 1.04\%$	$75.8 \pm 1.01\%$	$1.65 \pm 1.36\%$
Evo. (ours)	$\mathbf{1.51 \pm 0.26}\%$	$0.78 \pm 0.18\%$	$6.03 \pm 6.99\%$	$\mathbf{1.98 \pm 0.72}\%$	$0.83 \pm 0.15\%$
Causal Evo. (ours)	$2.12 \pm 0.67\%$	$\mathbf{0.75 \pm 0.12}\%$	$\mathbf{5.99 \pm 5.25}\%$	$2.28 \pm 0.76\%$	$\mathbf{0.71 \pm 0.007}\%$