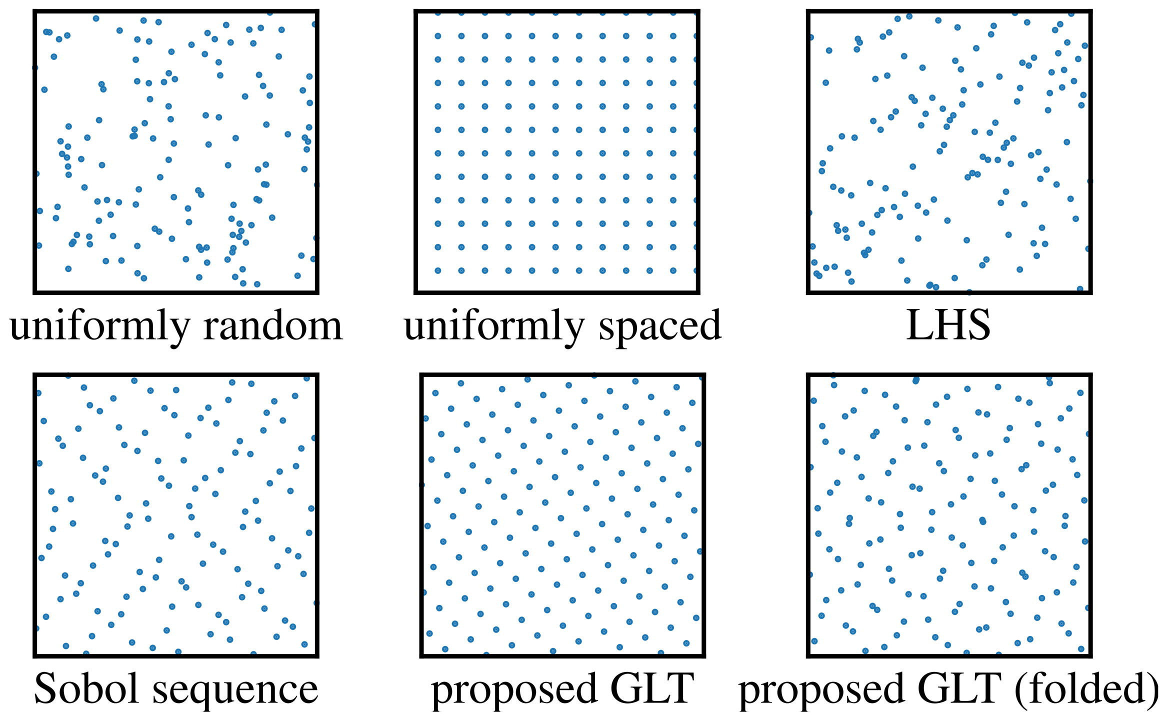
Number Theoretic Accelerated Learning of Physics-Informed Neural Networks

Takashi Matsubara (Hokkaido University), Takaharu Yaguchi (Kobe University)

AAAI-25

Summary

- Physics-informed neural networks (PINNs) solve partial differential equations (PDEs) by training neural networks as basis functions.
- PINNs approximate infinite-dimensional PDE solutions with finite collocation points.
- Minimizing discretization errors by selecting suitable points is essential for accelerating the learning process.
- Inspired by number theoretic methods for numerical analysis, we introduce good lattice training (GLT) and periodization tricks, which ensure the conditions required by the theory.
- Our experimental results demonstrate that GLT requires *2-7 times fewer collocation points*, resulting in lower computational cost, while achieving competitive performance.



Background and Theory

Definitions

- A PDE is expressed as $\mathcal{N}[u]=0$ for a (possibly nonlinear) differential operator $\mathcal{N}$ and an unknown function $u:\Omega\to\mathbb{R}$ on the domain $\Omega\subset\mathbb{R}^s$
- PINNs train a neural network $\tilde{u}$ by minimizing physics-informed loss,

$$\frac{1}{N}\sum_{j=0}^{N-1}\|\mathcal{N}[\tilde{u}](\boldsymbol{x}_j)\|^2=\frac{1}{N}\sum_{\boldsymbol{x}_j\in L^*}\|\mathcal{N}[\tilde{u}](\boldsymbol{x}_j)\|^2. \quad (1)$$

- (1) evaluates how the neural network $\tilde{u}$ satisfies the PDE $\mathcal{N}[\tilde{u}]=0$ at a finite set of N collocation points $\boldsymbol{x}_j$, L^*.
- However, the solutions u to PDEs are infinite-dimensional, and any distance involving $\tilde{u}$ or u needs to be defined by an integral over Ω.
- (1) serves as a finite approximation to the squared 2-norm,

$$|\mathcal{N}[\tilde{u}]\|_2^2=\int_{\boldsymbol{x}\in\Omega}\|\mathcal{N}[\tilde{u}](\boldsymbol{x})\|^2\mathrm{d}\boldsymbol{x} \quad (2)$$

on the function space $L^2(\Omega)$ for $\mathcal{N}[u]\in L^2(\Omega)$.
- Hence, the discretization errors should affect the training efficiency.

Error Estimation

- For simplicity, PDEs on $\Omega=[0,1]^s$ are considered.
- **Theorem 1.**
  - Suppose that the class of neural networks used for PINNs includes an ε_1-approximator $\tilde{u}_{\mathrm{opt}}$ to the exact solution u^* to the PDE $\mathcal{N}[u]=0$: $|u^*-\tilde{u}_{\mathrm{opt}}|\le\varepsilon_1$.
  - Suppose that (1) is an ε_2-approximation to (2) for the approximated solution $\tilde{u}$ and for $\tilde{u}_{\mathrm{opt}}$: $\left|\int_{[0,1]^s}\mathcal{N}[u](\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}-\frac{1}{N}\sum_{\boldsymbol{x}_j\in L^*}\mathcal{N}[u](\boldsymbol{x}_j)\right|\le\varepsilon_2$ for $u=\tilde{u}$ and $u=\tilde{u}_{\mathrm{opt}}$.
  - Suppose that there exist $c_{\mathrm{P}}>0$ and $c_{\mathrm{L}}>0$ such that

$$c_{\mathrm{P}}^{-1}\|u-v\|\le\|\mathcal{N}[u]-\mathcal{N}[v]\|\le c_{\mathrm{L}}\|u-v\|.$$

  - Then, $\|u^*-\tilde{u}\|\le(1+c_{\mathrm{P}}c_{\mathrm{L}})\varepsilon_1+c_{\mathrm{P}}\sqrt{\frac{1}{N}\sum_{\boldsymbol{x}_j\in L^*}\mathcal{N}[\tilde{u}](\boldsymbol{x}_j)+\varepsilon_2}$
- Theorem 1 suggests that the final error is composed of the error ε_1 depends on the network architecture and the discretization error ε_2.
- This study investigates a training method that easily gives small ε_2.

Reference

[1] Raissi et al., Physics-informed neural networks, JCP, 2019.
[2] Zeng et al., Competitive Physics Informed Networks, ICLR, 2023.
[3] Niederreiter, Random Number Generation and Quasi-Monte Carlo Methods, SIAM, 1992.
[4] Sloan and Joe, Lattice Methods for Multiple Integration, Clarendon Press, 1994.
[5] Zaremba, Applications of Number Theory to Numerical Analysis, Academic Press, 1972.

Method: Good Lattice Training

Derivations

- **Def. 2.** A lattice L in $\mathbb{R}^s$ is a finite set of points in $\mathbb{R}^s$ closed under addition and subtraction.
- The set of collocation points is $L^*\coloneqq\{\text{the decimal part of }\boldsymbol{x}|\boldsymbol{x}\in L\}$.
- Suppose $\varepsilon(\boldsymbol{x})\coloneqq\mathcal{N}[\tilde{u}](\boldsymbol{x})$ is periodic and smooth enough, admitting the Fourier series expansion: $\varepsilon(\boldsymbol{x})=\sum_{\boldsymbol{h}}\hat{\varepsilon}(\boldsymbol{h})\exp(2\pi\mathrm{i}\boldsymbol{h}\cdot\boldsymbol{x})$, where i denotes the imaginary unit and $\boldsymbol{h}=(h_1,h_2,\dots,h_s)\in\mathbb{Z}^s$.
- $|(2)-(1)|=\left|\frac{1}{N}\sum_{j=0}^{N-1}\sum_{\boldsymbol{h}\in\mathbb{Z}^s,\boldsymbol{h}\neq0}\hat{\varepsilon}(\boldsymbol{h})\exp(2\pi\mathrm{i}\boldsymbol{h}\cdot\boldsymbol{x}_j)\right|$ (3)
- **Def. 3.** A dual lattice $L^\top$ of L is $L^\top\coloneqq\{\boldsymbol{h}\in\mathbb{R}^s\,|\boldsymbol{h}\cdot\boldsymbol{x}\in\mathbb{Z},\forall x\in L\}$.
- **Lemma 4.** For $\boldsymbol{h}\in\mathbb{Z}^s$, $\frac{1}{N}\sum_{j=0}^{N-1}\hat{\varepsilon}(\boldsymbol{h})\exp(2\pi\mathrm{i}\boldsymbol{h}\cdot\boldsymbol{x}_j)=\begin{cases}1 & (\boldsymbol{h}\in\boldsymbol{L}\text{^T})\\ 0 & (\text{otherwise})\end{cases}$
- We restrict the lattice L to the form $\left\{\boldsymbol{x}\middle|\boldsymbol{x}=\frac{j}{N}\boldsymbol{z}\text{ for }j\in\mathbb{Z}\right\}$.
  - $L^\top=\{\boldsymbol{h}|\boldsymbol{h}\cdot\boldsymbol{z}\equiv0\ (\mathrm{mod}\ N)\}$
- $(3)\le\sum_{\boldsymbol{h}\in\mathbb{Z}^s,\boldsymbol{h}\neq0,\boldsymbol{h}\cdot\boldsymbol{z}\equiv0\ (\mathrm{mod}\ N)}|\hat{\varepsilon}(\boldsymbol{h})|$ (4)
- **Def. 5.** The Korobov space is $E_\alpha=\left\{f:[0,1]^s\to\mathbb{R}\middle|\exists c,|\hat{f}(\boldsymbol{h})|\le\frac{c}{(\bar{h}_1\bar{h}_2\dots\bar{h}_s)^\alpha}\right\}$, where $\hat{f}(\boldsymbol{h})$ is f's Fourier coefficients and $\bar{k}=\max(1,|k|)$ for $k\in\mathbb{R}$.
  - If α is an integer, for a function f to be in E^α, it is sufficient that f has continuous partial derivatives $\frac{\partial^{q_1+q_2+\dots+q_s}}{\partial_1^{q_1}\partial_2^{q_2}\dots\partial_s^{q_s}}f$, $0\le q_k\le\alpha\ (k=1,\dots,s)$.
- If $\mathcal{N}[\tilde{u}]$ belong to Korobov space,

$$(4)\le\sum_{\boldsymbol{h}\in\mathbb{Z}^s,\boldsymbol{h}\neq0,\boldsymbol{h}\cdot\boldsymbol{z}\equiv0\ (\mathrm{mod}\ N)}\frac{c}{(\bar{h}_1\bar{h}_2\dots\bar{h}_s)^\alpha} \quad (5)$$

- **Theorem 6.** For integers $N\ge2$ and $s\ge2$, there exists a $\boldsymbol{z}\in\mathbb{Z}^s$ such that $P_\alpha(\boldsymbol{z},N)\le\frac{(2\log N)^{\alpha s}}{N^\alpha}+O\left(\frac{(\log N)^{\alpha s-1}}{N^\alpha}\right)$ for $P_\alpha(\boldsymbol{z},N)=\frac{1}{(\bar{h}_1\bar{h}_2\dots\bar{h}_s)^\alpha}$.
- **Theorem 7.** Suppose that the activation function of $\tilde{u}$ and hence $\tilde{u}$ itself are sufficiently smooth so that there exists an $\alpha>0$ such that $\mathcal{N}[\tilde{u}]\in E^\alpha$. Then, for given integers $N\ge2$ and $s\ge2$, there exists an integer vector $\boldsymbol{z}\in\mathbb{Z}^s$ such that $L^*=\{\text{the decimal part of }\frac{j}{N}\boldsymbol{z}|\ j=0,\dots,N-1\}$ is a "*good lattice*" in the sense that

$$\left|\int_{\boldsymbol{x}\in\Omega}\|\mathcal{N}[\tilde{u}](\boldsymbol{x})\|^2\mathrm{d}\boldsymbol{x}-\frac{1}{N}\sum_{\boldsymbol{x}_j\in L^*}\|\mathcal{N}[\tilde{u}](\boldsymbol{x}_j)\|^2\right|=O\left(\frac{(\log N)^{\alpha s}}{N^\alpha}\right) \quad (6)$$

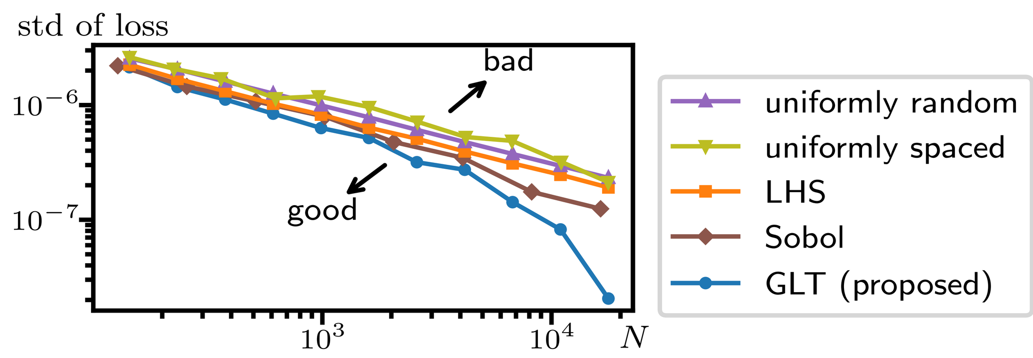
  - This rate is much better than that of the uniformly random sampling (i.e., the Monte Carlo method), which is of $O(1/N^{1/2})$.
- We call the training method that minimizes (1) for a lattice L satisfying (6) the *good lattice training (GLT).*

Periodization and Randomization Tricks

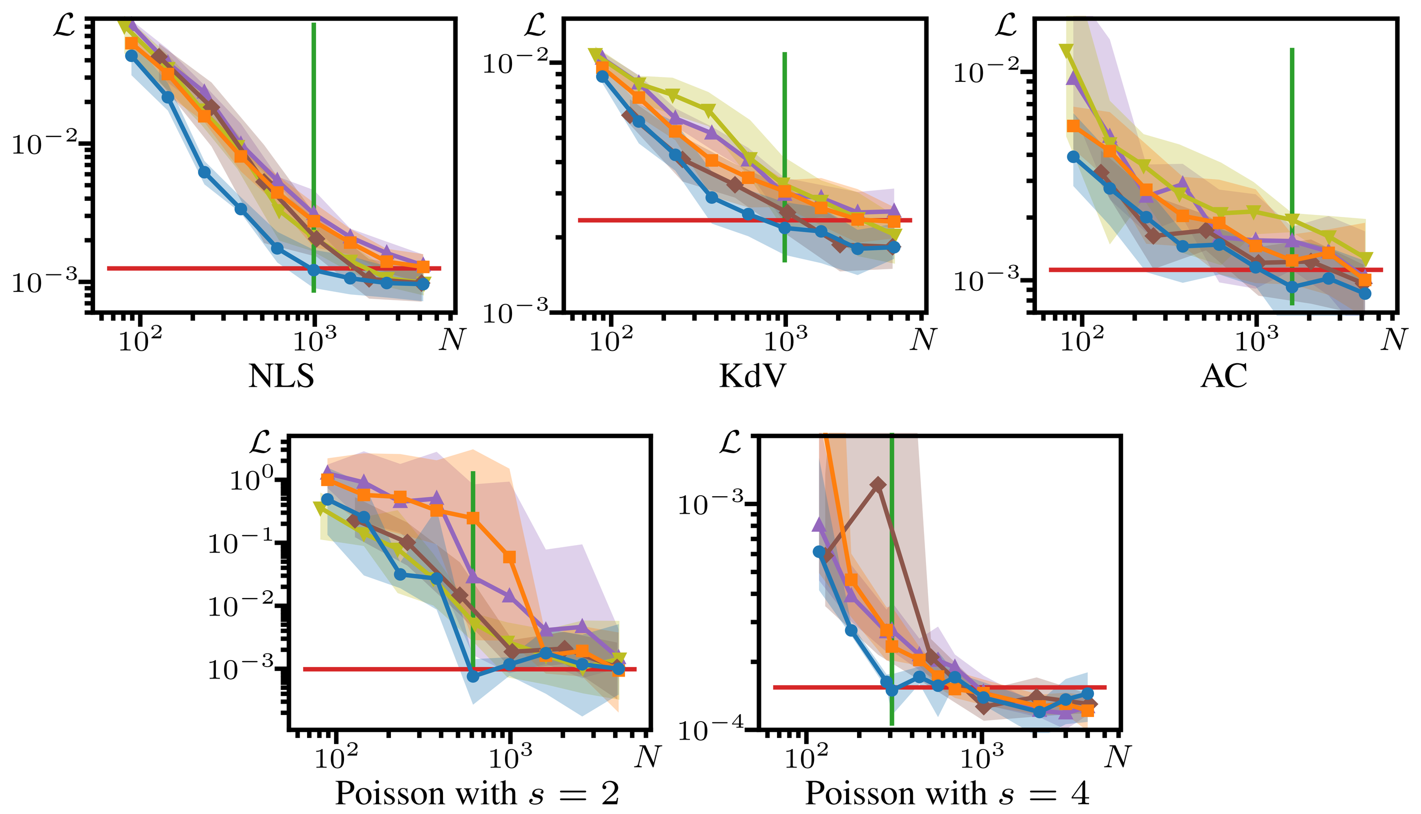
- We ensure the periodicity and boundary conditions as:
  - Given an initial condition, extend the lattice twice as much along the time coordinate and fold it.
  - Given a periodic boundary condition to the k-th axis, map the coordinate x_k to a unit circle in two-dimensional space.
  - Given a Dirichlet boundary condition $u=0$ to the k-th axis, treat the output $u(\dots,x_k,\dots)$ multiplied by $x_k(1-x_k)$ as the solution u.
- We randomly shift the collocation points at every training iteration, just like data augmentation.

Results: Competitive Performance with Much Fewer Collocation Points

- Std of (1) as an approximator to |(2)−(1)| because of $\mathbb{E}[(1)]=(2)$.

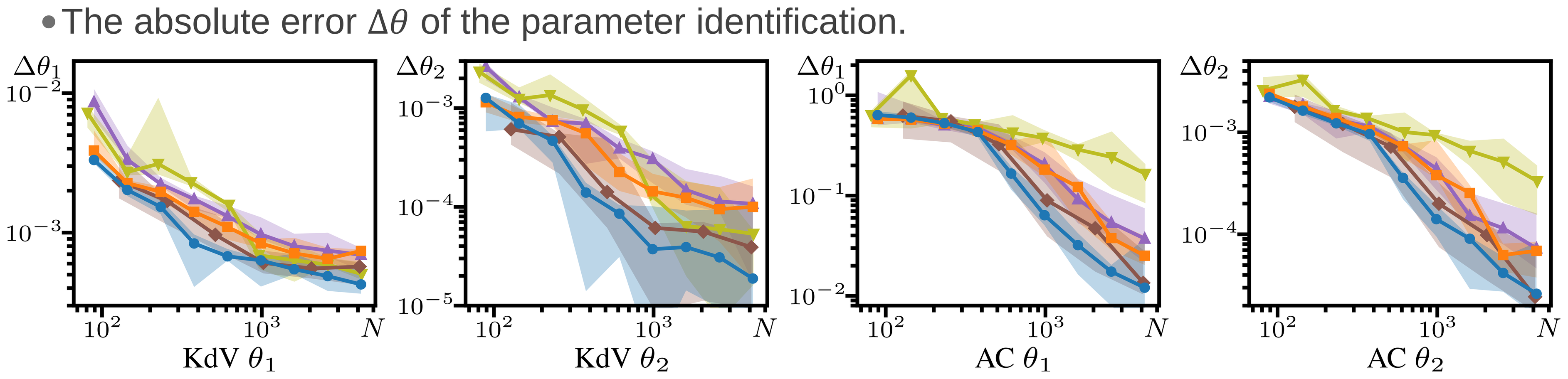


- The relative error, $\mathcal{L}[u;\tilde{u}]=\left(\sum_{\boldsymbol{x}\in L_e}|u-\tilde{u}|^2(\boldsymbol{x})\right)^{\frac{1}{2}}/\left(\sum_{\boldsymbol{x}\in L_e}|\tilde{u}|^2(\boldsymbol{x})\right)^{\frac{1}{2}}$, of u.

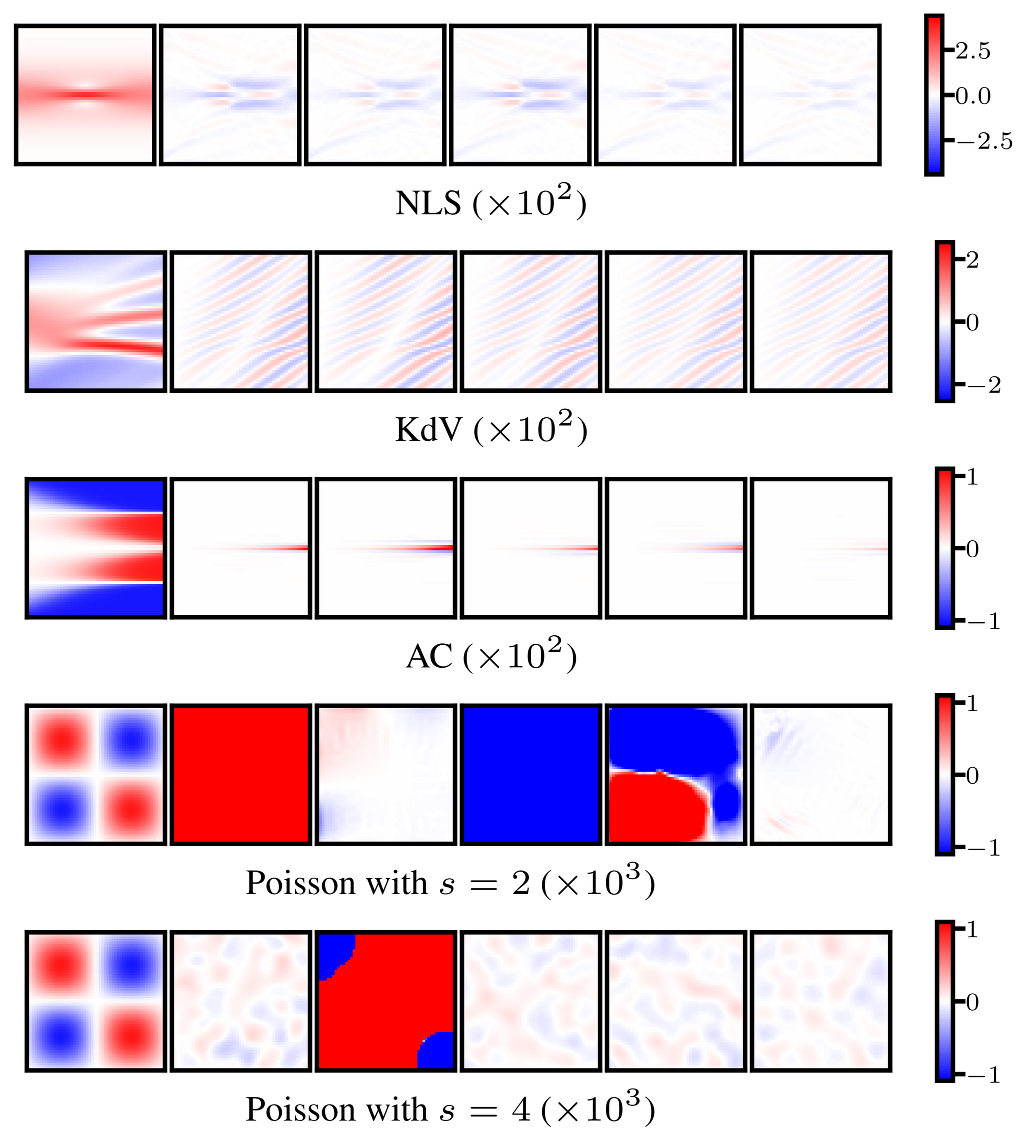


	# of points $N^\dagger$					relative error $\mathcal{L}^\ddagger$				
	NLS	KdV	AC	Poisson		NLS	KdV	AC	Poisson	
				$s=2$	$s=4$				$s=2$	$s=4$
▲ uniformly random	>4,181	>4,181	4,181	>4,181	1,019	3.11	2.97	1.55	28.53	0.28
▼ uniformly spaced	2,601	4,225	>4,225	>4,225	>4,096	2.15	3.28	1.95	5.16	1437.12
■ LHS	>4,181	4,181	4,181	4,181	701	2.75	3.06	1.25	246.29	0.24
◆ Sobol	2,048	2,048	4,096	>4,096	1,024	2.05	2.52	1.22	14.74	1.22
● GLT (proposed)	**987**	**987**	**1,597**	**610**	**307**	**1.22**	**2.19**	**0.93**	**0.76**	**0.15**

- The absolute error $\Delta\theta$ of the parameter identification.



- Visualization of $\tilde{u}$ and $|u-\tilde{u}|$.
  - Ground truth u, and errors $\tilde{u}-u$ for uniformly random, uniformly spaced, LHS, Sobol, and GLT.



- Competitive PINNs