Number Theoretic Accelerated Learning of Physics-Informed Neural Networks

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

Background

Partial Differential Equations (PDEs)

- appear in physical simulations,
  - such as weather forecasting and molecular dynamics simulations,
- include the Navier-stokes equation and the nonlinear Schrodinger equation.
- are defined 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 $\Omega \subset \mathbb{R}^s$.
- have been solved by various numerical methods,
  - including finite difference methods, finite volume methods, and spectral methods.

(Furihata and Matsuo 2010; Morton and Mayers 2005; Thomas 1995)

Background

Physics-informed neural networks (PINNs)

- were proposed an alternative numerical method,
  potentially computationally efficient and generally applicable.
- train a neural network $\tilde{u}\colon \Omega \to \mathbb{R}$ to represent the solution u.

Namely, PINNs

- minimize the *physics-informed loss*

$$\frac{1}{N}\sum_{j=0}^{N-1}\left\|\mathcal{N}[\tilde{u}]\left(\boldsymbol{x}_j\right)\right\|^2$$

  at a finite set of collocation points, $\left\{\boldsymbol{x}_j\right\}_j$.
- encourage the output $\tilde{u}$ to satisfy the PDE $\mathcal{N}[\tilde{u}](\boldsymbol{x}_j) = 0$.

(Raissi, Perdikaris, and Karniadakis 2019)

Problem

The solutions u to PDEs are inherently infinite-dimensional.

- Distances for the output $\tilde{u}$ or the solution u need to be defined by an integral over the domain Ω

In this regards,

- the physics-informed loss serves as a finite approximation to the squared 2-norm

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

  on the function space $L^2(\Omega)$ for $\mathcal{N}[u] \in L^2(\Omega)$.
- *the discretization errors should affect the training efficiency.*
  - A smaller number N leads to an inaccurate approximation.
  - A larger number N increases the computational cost.

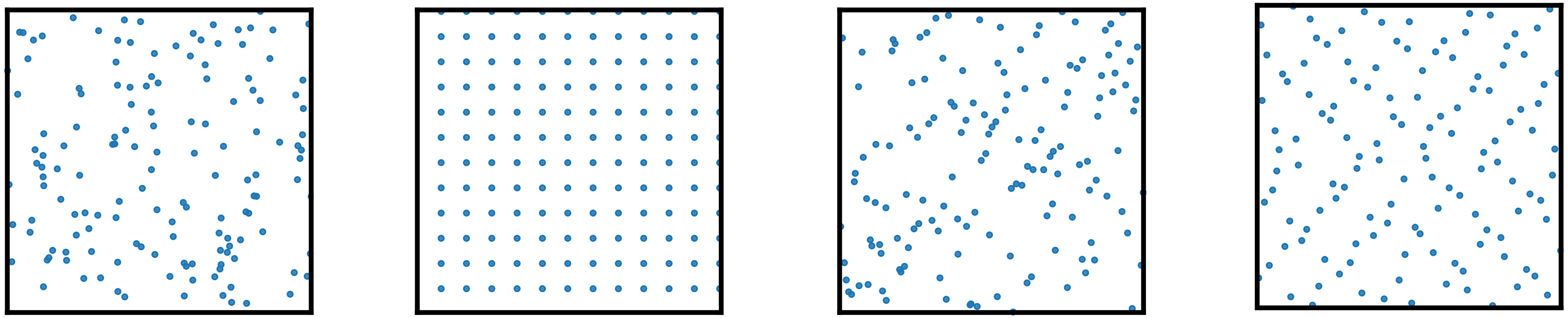
(Bihlo and Popovych 2022; Sharma and Shankar 2022)

Related Work

Sampling methods include (Jin et al. 2021; Krishnapriyan et al. 2022)

- uniformly random sampling (i.e., the Monte Carlo method)
- uniformly spaced sampling (Wang, Teng, and Perdikaris 2021; Wang, Yu, and Perdikaris 2022).
- Latin hypercube sampling (LHS) (Raissi, Perdikaris, and Karniadakis, 2019; Zeng et al. ,2023)
- Sobol sequence (a quasi-Monte Carlo method.)

(Lye, Mishra, and Ray 2020; Longo et al. 2021; Mishra and Molinaro 2021).



uniformly random | uniformly spaced | LHS | Sobol sequence

A more efficient way to sample collocation points?

Theory

Preliminary

- The domain $\Omega = [0,1]^s$
- A set of collocation points $L^* = \{\boldsymbol{x}_j \mid j = 0, \ldots, N-1\}$
- Physics-informed loss and its variants

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

  - where $\mathcal{P}[\tilde{u}](\boldsymbol{x}) = \|\mathcal{N}[\tilde{u}](\boldsymbol{x})\|^2$ for the original PINNs, and $\mathcal{P}[\tilde{u}](\boldsymbol{x}) = D(\boldsymbol{x})\mathcal{N}[\tilde{u}](\boldsymbol{x})$ for CPINNs (Zeng et al. 2023).
- The integral loss

$$\int_{\boldsymbol{x}\in\Omega} \mathcal{P}[\tilde{u}](\boldsymbol{x})\mathrm{d}\boldsymbol{x} \quad (2)$$

- The practical minimization of (1) essentially minimizes the approximation of (2).

Theory

Theorem 1

- Suppose that the class of neural networks includes an ε_1-approximator $\tilde{u}_{opt}$ to the exact solution u^* to the PDE $\mathcal{N}[u] = 0$: $\|u^* - \tilde{u}_{opt}\| \leq \varepsilon_1$.
- Suppose that (1) is an ε_2-approximation of (2) for a neural nework $\tilde{u}$ and for $\tilde{u}_{opt}$: $|(2) - (1)| < \varepsilon_2$ for $u = \tilde{u}$ and $u = \tilde{u}_{opt}$.
- Suppose also that there exist $c_p > 0$ and $c_L > 0$ such that $c_p^{-1} \|u - v\| \leq \|\mathcal{N}[u] - \mathcal{N}[v]\| \leq c_L \|u - v\|$.
- Then, $\|u^* - \tilde{u}\| \leq \underbrace{(1 + c_p c_L)\varepsilon_1}_{\text{network capacity}} + \underbrace{c_p\sqrt{(1) + \varepsilon_2}}_{\text{discretization error}}$.

How to reduce the second term?

Good Lattice Training

We propose Good Lattice Training (GLT)

- We use a number theoretic numerical analysis to accelerate the training of PINNs.
- Suppose $\Omega = [0, 1]^s$, the loss $\mathcal{P}[\tilde{u}]$ be periodic on $\mathbb{R}^s$ with a period of 1.

Definition 2

- A lattice L in $\mathbb{R}^s$ is defined as a finite set of points in $\mathbb{R}^s$ that is closed under addition and subtraction.

Then,

- the set of collocation points is defined as

$$L^* = \{\boldsymbol{x}_j \mid j = 0, \ldots, N-1\} := \{the\ decimal\ part\ of\ \boldsymbol{x} | \boldsymbol{x} \in L\} \in [0, 1]^s.$$

(Niederreiter 1992; Sloan and Joe 1994; Zaremba 1972)

Good Lattice Training

- Suppose $\varepsilon(\boldsymbol{x}) \coloneqq \mathcal{P}[\tilde{u}](\boldsymbol{x})$ is smooth enough, admitting the Fourier series expansion:

$$\varepsilon(\boldsymbol{x}) \coloneqq \mathcal{P}[\tilde{u}](\boldsymbol{x}) = \sum_{\boldsymbol{h}} \hat{\varepsilon}(\boldsymbol{h}) \exp(2\pi \mathrm{i} \boldsymbol{h} \cdot \boldsymbol{x})$$

- Then,

$$|(2) - (1)| = \left| \frac{1}{N} \sum_{j=0}^{N-1} \sum_{\boldsymbol{h} \in \mathbb{Z}^s, \boldsymbol{h} \neq 0} \hat{\varepsilon}(\boldsymbol{h}) \exp\left(2\pi \mathrm{i} \boldsymbol{h} \cdot \boldsymbol{x}_j\right) \right| \quad (3)$$

  - because the Fourier mode of $\boldsymbol{h} = 0$ is equal to the integral $\int_{[0,1]^s} \varepsilon(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$.

Definition 3

- A dual lattice $L^\top$ of a lattice L is defined as $L^\top \coloneqq \{\boldsymbol{h} \in \mathbb{R}^s | \boldsymbol{h} \cdot \boldsymbol{x} \in \mathbb{Z}, \forall \boldsymbol{x} \in L\}$.

Lemma 4

- For $\boldsymbol{h} \in \mathbb{Z}^s$, it holds that $\frac{1}{N} \sum_{j=0}^{N-1} \exp(2\pi \mathrm{i} \boldsymbol{h} \cdot \boldsymbol{x}) = 1$ if $\boldsymbol{h} \in L^\top$ and 0 otherwise

(Niederreiter 1992; Sloan and Joe 1994; Zaremba 1972)

Good Lattice Training

Good Lattice Training (GLT)

- We restrict the lattice L to the form $\left\{ \boldsymbol{x} \,\middle|\, \boldsymbol{x} = \frac{j}{N}\boldsymbol{z} \text{ for } j \in \mathbb{Z} \right\}$ for an integer vector $\boldsymbol{z}$.
  - The set L^* of collocation points is $\left\{ \textit{the decimal part of } \frac{j}{N}\boldsymbol{z} \,\middle|\, j = 0, \ldots, N-1 \right\}$.
  - $L^\top = \{\boldsymbol{h} \mid \boldsymbol{h} \cdot \boldsymbol{z} \equiv 0 \pmod N\}$.
- Then, $(3) \leq \sum_{\boldsymbol{h} \in \mathbb{Z}^s, \boldsymbol{h} \neq 0, \boldsymbol{h} \cdot \boldsymbol{z} \equiv 0 \pmod N\}} |\hat{\varepsilon}(\boldsymbol{h})|$ (4)
- To minimize $|(2) - (1)|$, find an integer vector $\boldsymbol{h}$ that minimizes (4).
  - In this sense, this is a number theoretic problem.

(Niederreiter 1992; Sloan and Joe 1994; Zaremba 1972)

Good Lattice Training

Definition 5

- The function space that is defined as

$$E_\alpha = \left\{ f : [0,1]^s \to \mathbb{R} \,\middle|\, \exists c, \left|\hat{f}(\boldsymbol{h})\right| \le \frac{c}{(\bar{h}_1 \bar{h}_2 \cdots \bar{h}_s)^\alpha} \right\}$$

  is called the *Korobov space*,

  - where $\hat{f}(\boldsymbol{h})$ is the Fourier coefficients of f and $\bar{k} = \max(1, |k|)$ for $k \in \mathbb{R}$.
  - For example, if a function $f(x, y): \mathbb{R}^2 \to \mathbb{R}$ has continuous f_x, f_y, f_{xy}, then $f \in E_1$.

(Niederreiter 1992; Sloan and Joe 1994; Zaremba 1972)

Good Lattice Training

- Hence, if $\mathcal{P}[\tilde{u}]$ and the neural network belong to Koborov space,

$$(4) \leq \sum_{\boldsymbol{h}\in\mathbb{Z}^s, \boldsymbol{h}\neq 0, \boldsymbol{h}\cdot\boldsymbol{z} \equiv 0 \ (\mathrm{mod}\ N)\}} \frac{c}{(\bar{h}_1 \bar{h}_2 \cdots \bar{h}_s)^{\alpha}} \tag{5}$$

Theorem 6

- For integers $N \geq 2$ and $s \geq 2$, there exists a $\boldsymbol{z} \in \mathbb{Z}^s$ such that

$$P_\alpha(\boldsymbol{z}, N) \leq \frac{(2 \log N)^{\alpha s}}{N^\alpha} + O\left(\frac{(\log N)^{\alpha s - 1}}{N^\alpha}\right) \text{ for } P_\alpha(\boldsymbol{z}, N) = \frac{1}{(\bar{h}_1 \bar{h}_2 \cdots \bar{h}_s)^{\alpha}}$$

- Values of $\boldsymbol{z}$ have been explored in the field of number theoretic numerical analysis.
  - Successive Fibonacci numbers are theoretically optimal for $s = 2$.
  - Numerical tables are available for $s > 2$. (Fang and Wang, 1994); Keng and Yuan, 1981))
  - Some algorithms find optimal values at the computational cost of $O(N^2)$.

Good Lattice Training

Theorem 7 (main result)

- 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 $\mathcal{P}[\tilde{u}] \in E_\alpha$.
- Then, for given integers $N \geq 2$ and $s \geq 2$, there exists an integer vector $\boldsymbol{z} \in \mathbb{Z}^s$ such that $L^* = \{the\ decimal\ part\ of\ \frac{j}{N}\boldsymbol{z} \mid j = 0, \ldots, N-1\}$ is a "good lattice" in the sense that

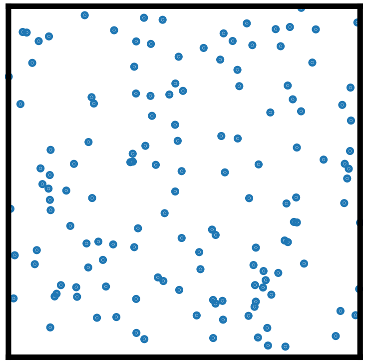
$$\left|\int_{\boldsymbol{x}\in\Omega} \mathcal{P}[\tilde{u}](\boldsymbol{x})\mathrm{d}\boldsymbol{x} - \frac{1}{N}\sum_{\boldsymbol{x}\in L^*} \mathcal{P}[\tilde{u}](\boldsymbol{x})\right| = O\left(\frac{(\log N)^{\alpha s}}{N^\alpha}\right) \quad (6)$$

Comparison

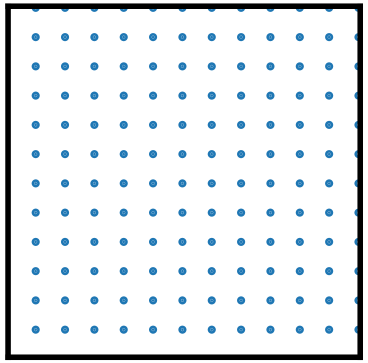
- This rate is better than the of the Monte Carlo method, which is of $O\left(N^{-1/2}\right)$ if $\alpha \geq 1$.

Good Lattice Training

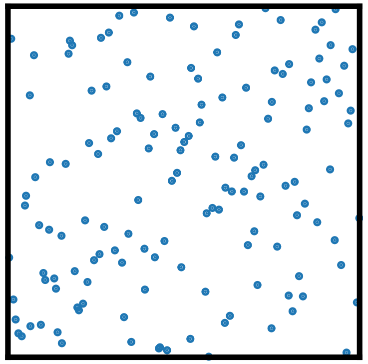
Comparison



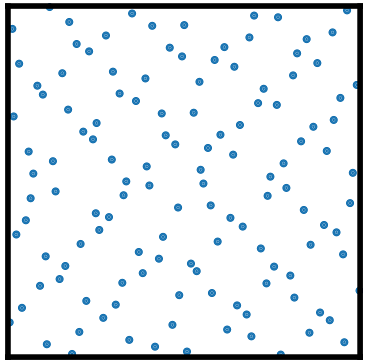
uniformly random



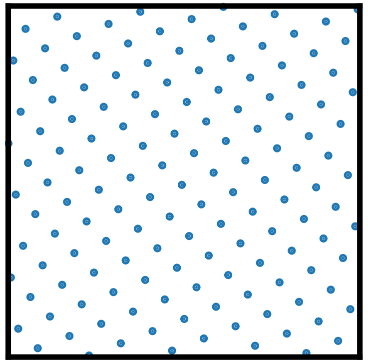
uniformly spaced



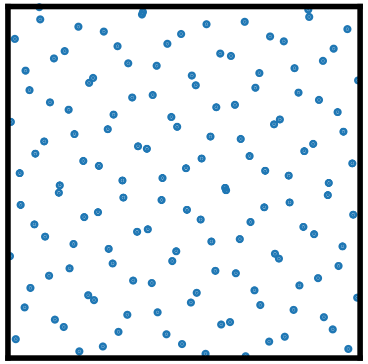
LHS



Sobol sequence



proposed GLT



proposed GLT (folded)

Periodization & Randomization Tricks

Good Lattice Training (GLT)

- suppose the periodicity and smoothness of the solutions and neural networks.

Periodization Trick

- ensures the periodicity
  - by folding the time coordinate
    ⇨i.e., using $\hat{t}$ satisfysing $t = 2\hat{t}$ if $\hat{t} < 0.5$ and $t = 2(1 - \hat{t})$ otherwise.
  - by projecting the space coordinate to a unit circle in a 2D space
    for the periodic boundary condition.
  - by multiplying $x(1 - x)$
    for the Dirichlet boundary condition $u = 0$ at $\partial\Omega$.

Periodization & Randomization Tricks

Good Lattice Training (GLT)

- suppose the periodicity and smoothness of the solutions and neural networks.

Randomization Trick

- $L^* = \{the\ decimal\ part\ of\ \frac{j}{N}\boldsymbol{z} + \boldsymbol{r} \mid j = 0, \ldots, N-1\}$
  - where $\boldsymbol{r}$ follows the uniform distribution over the unit cube $[0,1]^s$.
- When using the stochastic gradient descent (SGD) algorithm,
  resampling the random numbers $\boldsymbol{r}$ at each training iteration
  prevents the neural network from overfitting and improves training efficiency.

Experiments

Experiments

- The Adam for 20,000 iterations
  - better than the L-BFGS-B method preceded by the Adam for 50,000 iterations.

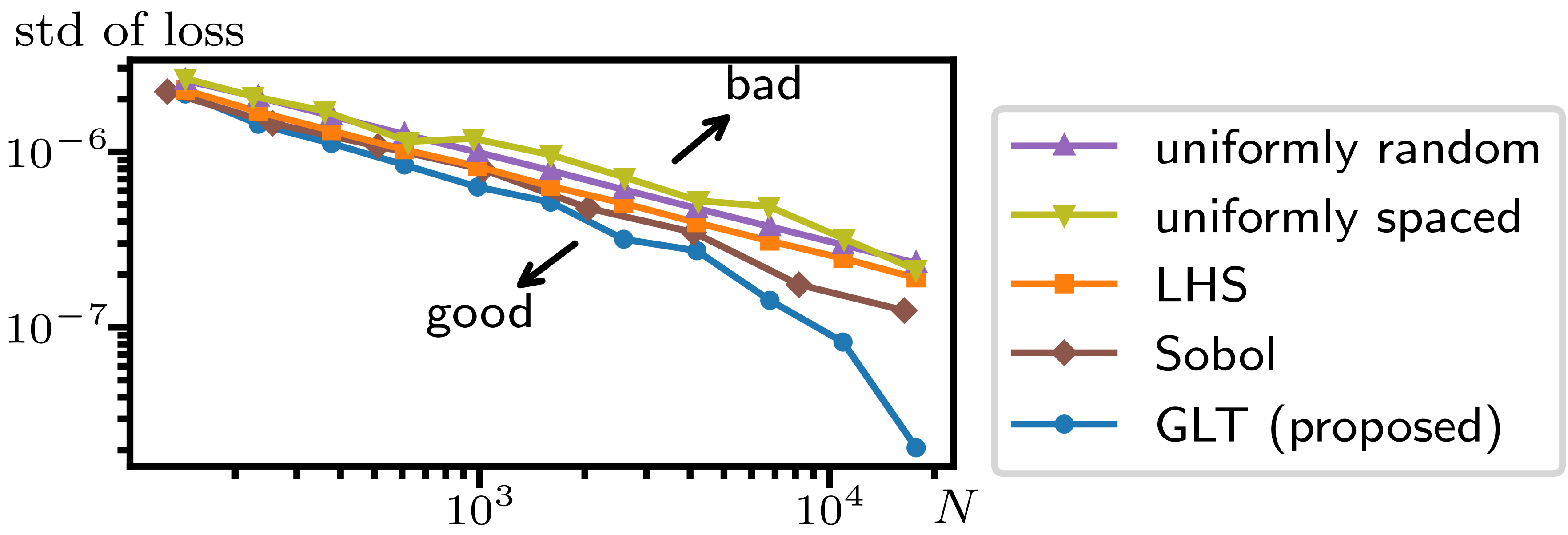
Evaluation

- The relative error $\mathcal{L}(\tilde{u}, u) = \frac{\sqrt{\sum_{j=0}^{N-1} \left\| \tilde{u}(\boldsymbol{x}_j) - u(\boldsymbol{x}_j) \right\|^2}}{\sqrt{\sum_{j=0}^{N-1} \left\| u(\boldsymbol{x}_j) \right\|^2}} \simeq \frac{|\tilde{u} - u|_2}{|u|_2}$

Results: Discretization error

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

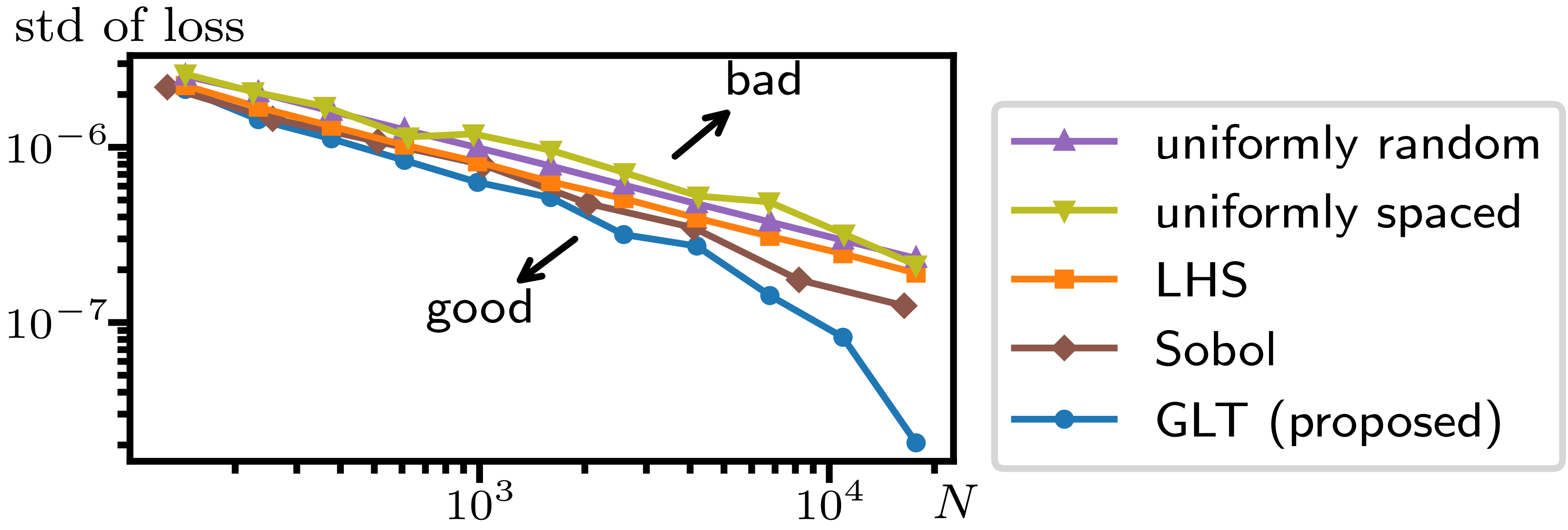
- After training the PINNs on the 2D NLS equation with $N = 610$ collocation points determined by LHS.



Results: Discretization error

Convergence rate

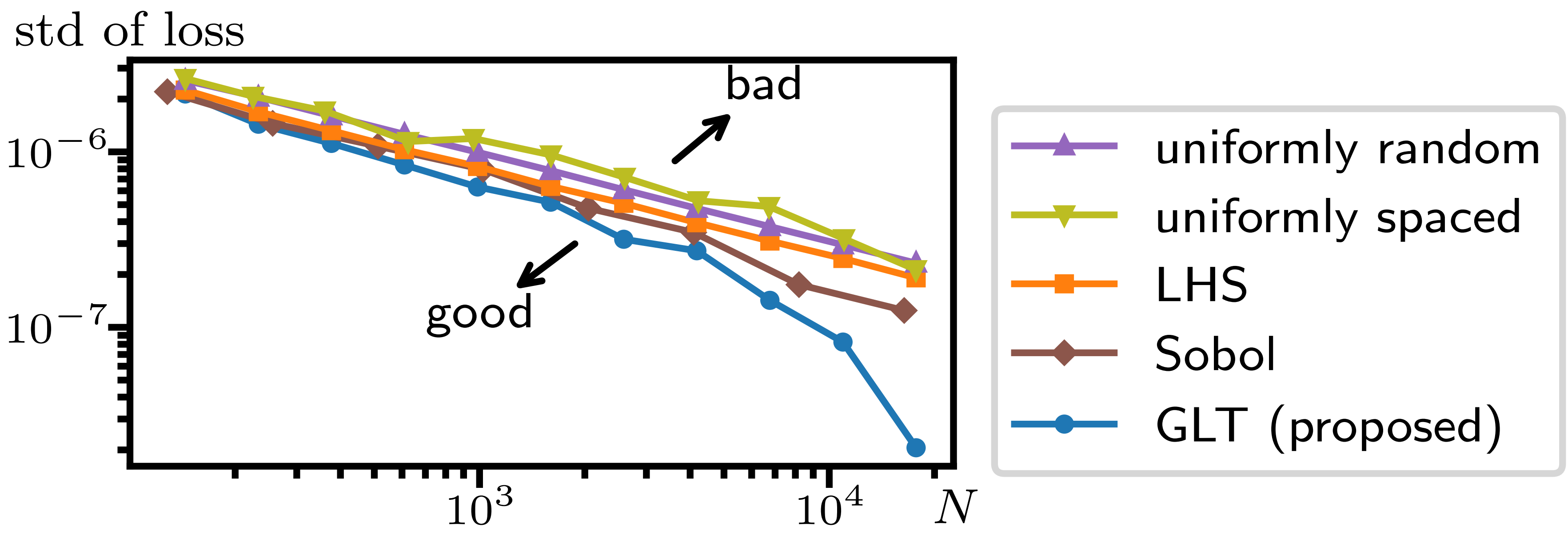
- This trend aligns with the theory: $O(N^{-1/2})$ for the uniformly random, $O\left(N^{-1/s}\right)$ for the uniformly spaced, and $O\left(\frac{(\log N)^s}{N}\right)$ for the Sobol sequence.



Results: Discretization error

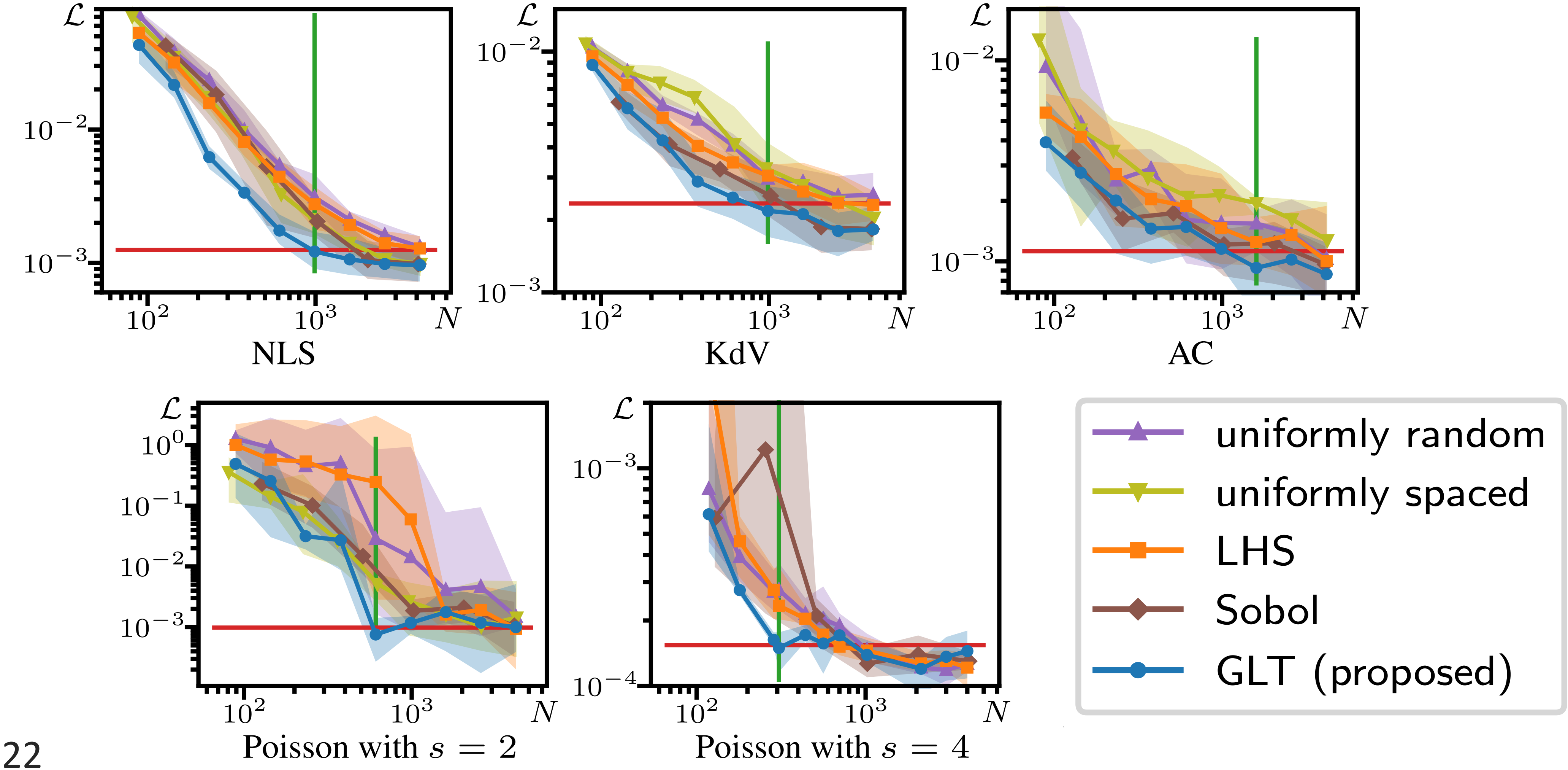
GLT

- demonstrates a further accelerated reduction as the number N increases.
- is comparable with Sobol sequence if $\alpha = 1$ and much better if $\alpha > 1$.
- does not require the hyperparameter adjustment depending on α.



Results: Relative Error

The relative error converges the fastest with GLT.



Theory: Recall

Theorem 1

- Suppose that the class of neural networks includes an ε_1-approximator $\tilde{u}_{opt}$ to the exact solution u^* to the PDE $\mathcal{N}[u] = 0$: $\|u^* - \tilde{u}_{opt}\| \leq \varepsilon_1$.
- Suppose that (1) is an ε_2-approximation of (2) for the approximated solution $\tilde{u}$ and also for $\tilde{u}_{opt}$: $|(2) - (1)| < \varepsilon_2$ for $u = u^*$ and $u = \tilde{u}_{opt}$.
- Suppose also that there exist $c_p > 0$ and $c_L > 0$ such that $c_p^{-1} \|u - v\| \leq \|\mathcal{N}[u] - \mathcal{N}[v]\|$.
- Then, $\|u^* - \tilde{u}\| \leq \underbrace{(1 + c_p c_L)\varepsilon_1}_{\text{network capacity}} + \underbrace{c_p\sqrt{(1) + \varepsilon_2}}_{\text{discretization error}}$.

The convergence is due to the network capacity.

Results: Relative Error

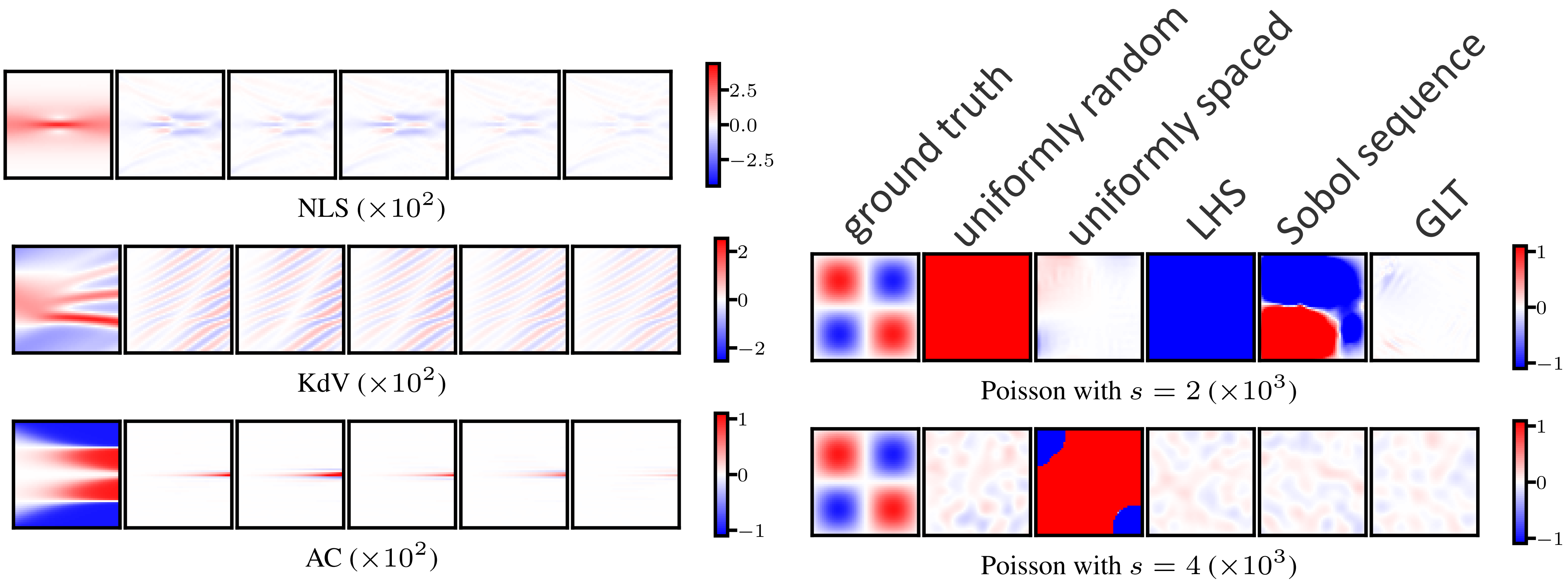
x 2-7 speed up.

- The same errors with the smaller numbers of collocation points.
- The smaller errors with the same numbers of collocation points.

	# 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**

Results: Visualization

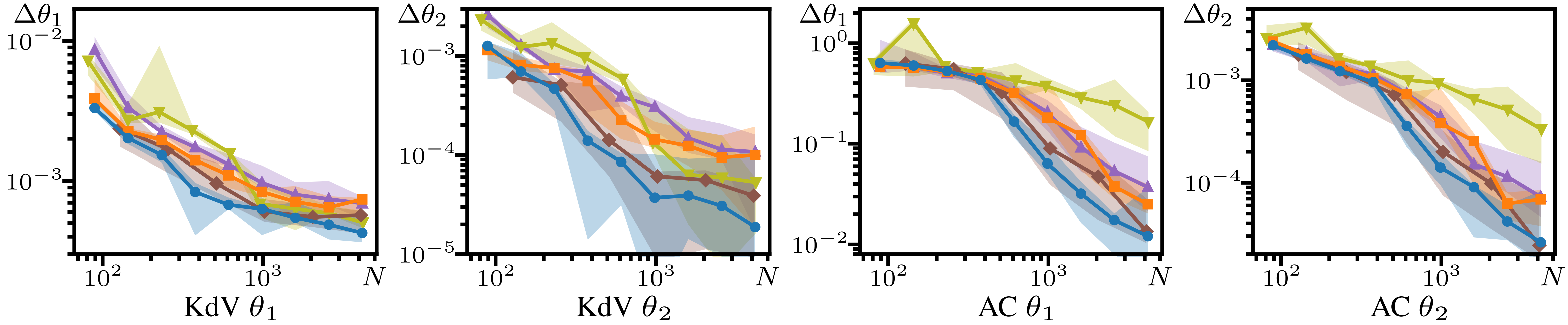
Errors with the same number of colocation points.



Results: Parameter Identification

Errors in the parameter identification

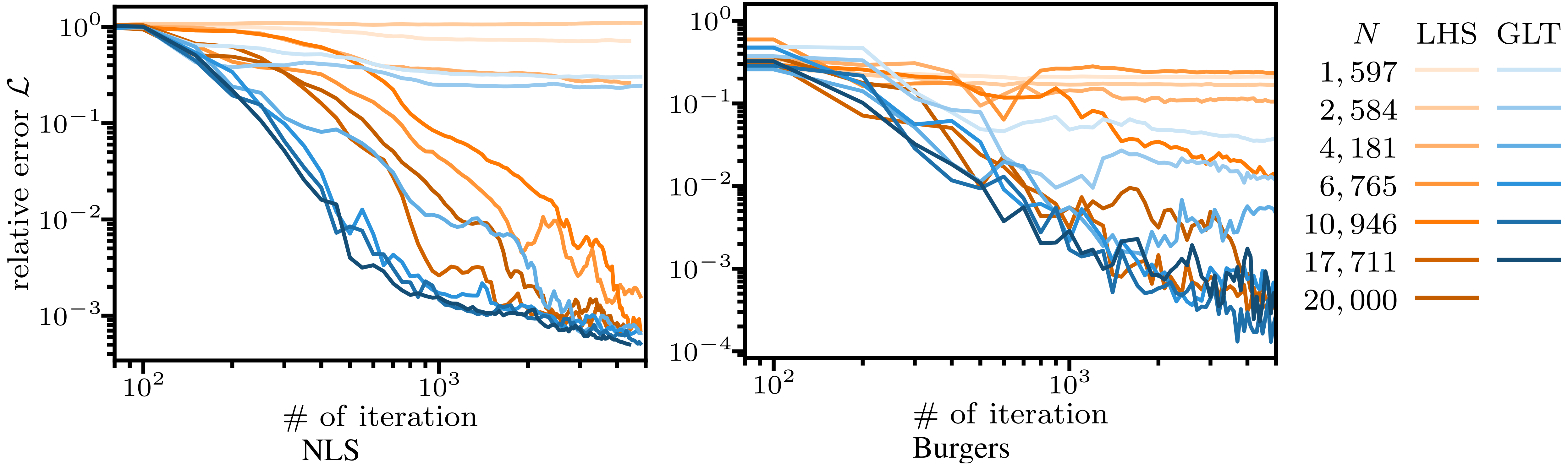
- Smaller errors with fewer collocation points.



Results: CPINNs

Fast convergence of the relative errors with fewer collocation points

- x 2-4 speed up.



Conclusion

We propose good lattice training

- Number theoretic numerical analysis method accelerates the training of PINNs.
  - by reducing the number of collocation points to 1/7-1/2.
- Periodization and randomization tricks ensure the conditions required by the theory.
  - Without these tricks, the performance significantly degraded.
- GLT worked well also for PINNs variants (namely, CPINNs).

	relative error $\mathcal{L}^{\ddagger}$		
	NLS	KdV	AC
▲ uniformly random	3.18	17.30	382.51
▼ uniformly spaced	1.98	16.08	94.33
■ LHS	2.78	15.14	158.71
◆ Sobol	2.21	13.28	94.35
● GLT	**1.31**	**12.30**	**84.50**
● GLT (with tricks)	**1.22**	**2.19**	**0.93**