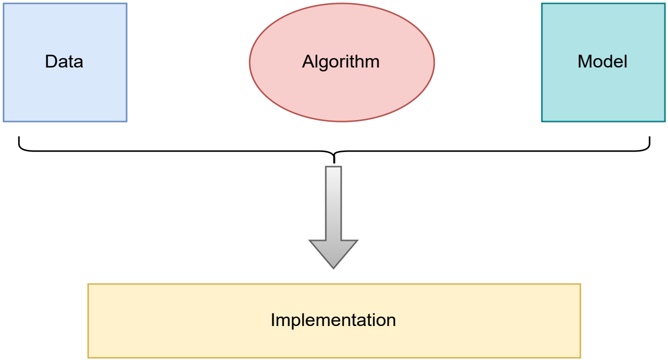
Adversarial Adaptive Sampling: Unify PINN and Optimal Transport for the Approximation of PDEs

PKU-Changsha Institute for Computing and Digital Economy

Kejun Tang
tangkejun@icode.pku.edu.cn

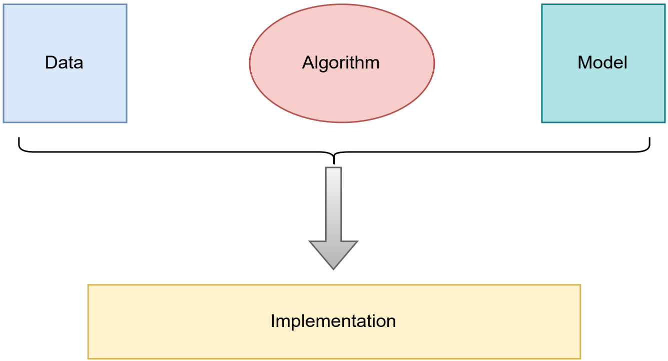
Joint work with Jiayu Zhai (Co-first Author, ShanghaiTech University), Xiaoliang Wan (Louisiana State University) and Chao Yang (Peking University)

Big data era: data-driven



- Model: deep neural networks, physical models, or coupling ...
- Data: labeled, unlabled, random samples
- Algorithm: various optimization methods

Big data era: data-driven



data is oil

- model is driven by data
- data has the influence on generalization

Deep learning for PDEs

$$\mathcal{L}\left(\mathbf{x}; u\left(\mathbf{x}\right)\right) = s(\mathbf{x}) \qquad \forall \mathbf{x} \in \Omega,$$
$$\mathfrak{b}\left(\mathbf{x}; u\left(\mathbf{x}\right)\right) = g(\mathbf{x}) \qquad \forall \mathbf{x} \in \partial\Omega.$$

$\mathcal{L}$: partial differential operator, $\mathfrak{b}$: boundary operator.

FEM:

1. **mesh**
2. **basis**

Deep methods:

1. **samples**
2. **neural networks**

Why deep methods

- fast inference
- tackle high dimensional problems

Deep learning for PDEs

$$\mathcal{L}(x; u(x)) = s(x) \qquad \forall (x) \in \Omega,$$
$$\mathfrak{b}(x; u(x)) = g(x) \qquad \forall (x) \in \partial\Omega.$$

$\mathcal{L}$: partial differential operator, $\mathfrak{b}$: boundary operator.

How deep methods do: a deep nets $u(\mathbf{x}; \Theta) \rightarrow u(\mathbf{x})$

$$\mathcal{J}(u(\mathbf{x}; \Theta)) = \|r(\mathbf{x}; \Theta)\|_{2,\Omega}^2 + \gamma \|b(\mathbf{x}; \Theta)\|_{2,\partial\Omega}^2,$$

where $r(\mathbf{x}; \Theta) = \mathcal{L}u(\mathbf{x}; \Theta) - s(\mathbf{x})$, $b(\mathbf{x}; \Theta) = \mathfrak{b}u(\mathbf{x}; \Theta) - g(\mathbf{x})$, and

$$\|r(\mathbf{x}; \Theta)\|_{2,\Omega}^2 = \int_\Omega r^2(\mathbf{x}; \Theta) d\mathbf{x}$$

An optimization problem: $\min_{\Theta} \mathcal{J}(u(\mathbf{x}; \Theta))$

Deep learning for PDEs

$$\begin{aligned} \mathcal{L}\left(x; u\left(x\right)\right) &= s(x) \qquad \forall \left(x, \right) \in \Omega, \\ \mathfrak{b}\left(x; ; u\left(x, \right)\right) &= g(x) \qquad \forall \left(x, \right) \in \partial\Omega. \end{aligned}$$

$\mathcal{L}$: partial differential operator, $\mathfrak{b}$: boundary operator.

How deep methods do: a deep nets $u(\mathbf{x}; \Theta) \to u(\mathbf{x})$

$$\mathcal{J}_N\left(u(\mathbf{x}; \Theta)\right) = \frac{1}{N_r} \sum_{i=1}^{N_r} r^2(\mathbf{x}_\Omega^{(i)}; \Theta) + \hat{\gamma} \frac{1}{N_b} \sum_{i=1}^{N_b} b^2(\mathbf{x}_{\partial\Omega}^{(i)}; \Theta),$$

$\mathbf{x}_\Omega^{(i)}$ drawn from Ω and $\mathbf{x}_{\partial\Omega}^{(i)}$ drawn from $\partial\Omega$

Key point: $\min_\Theta \mathcal{J}\left(u(\mathbf{x}; \Theta)\right) \to \min_\Theta \mathcal{J}_N\left(u(\mathbf{x}; \Theta)\right)$ discretize loss by uniform sampling or other quasi-random methods based on uniform samples

Deep learning for PDEs

$$u(\mathbf{x};\Theta^*) = \arg\min_{\Theta} J(u(\mathbf{x};\Theta)),$$

$$u(\mathbf{x};\Theta_N^*) = \arg\min_{\Theta} J_N(u(\mathbf{x};\Theta)).$$

$$\mathbb{E}\left(\|u(\mathbf{x};\Theta_N^*) - u(\mathbf{x})\|_\Omega\right) \leq \underbrace{\mathbb{E}\left(\|u(\mathbf{x},\Theta_N^*) - u(\mathbf{x};\Theta^*)\|_\Omega\right)}_{\text{statistical error}} + \underbrace{\|u(\mathbf{x};\Theta^*) - u(\mathbf{x})\|_\Omega}_{\text{approximation error}}$$

Our work: focus on how to reduce the statistical error

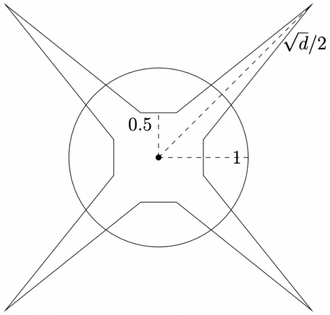
the capability of neural networks $\rightarrow$ approximation error

the strategy of loss discretization $\rightarrow$ statistical error

Key point: how to sample?

Geometric properties of high-dimensional spaces

uniformly distributed points in high-dimensional spaces

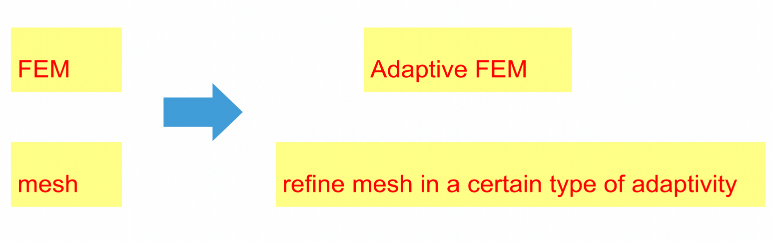


Most of the volume of a high-dimensional cube is located around its corner [Vershynin, High-Dimensional Probability, 2020]. Cube: $[-1,1]^d$

$$\mathbb{P}(\|\mathbf{x}\|_2^2 \leq 1) \leq \exp(-\frac{d}{10}).$$

Adaptivity

Question: is uniform sampling optimal for deep methods?



Observation:

1. uniform mesh is not optimal for FEM
2. choosing uniform samples is not a good choice for high-dimensional problems

Deep methods

lack of adaptivity $\rightarrow$ develop adaptive schemes

Related work of adaptive sampling

- RAR [Lu et. al, 2021]
- MCMC [Gao & Wang, 2023]
- DAS [Tang, Wan, and Yang, 2022]
- Gaussian mixture models [Gao et.al, 2023; Jiao et. al, 2023]
- . . .

Goal

- formulate two essential components, minimizing the residual and seeking the optimal training set, into one min-max objective functional

Adversarial adaptive sampling

Estimate the residual

$$\int_\Omega r^2(\mathbf{x};\Theta)d\mathbf{x} \approx \frac{1}{N_r}\sum_{i=1}^{N_r} r^2(\mathbf{x}_\Omega^{(i)};\Theta),$$

key point

- reduce the variance of r^2
- the profile of the residual needs to be nearly uniform

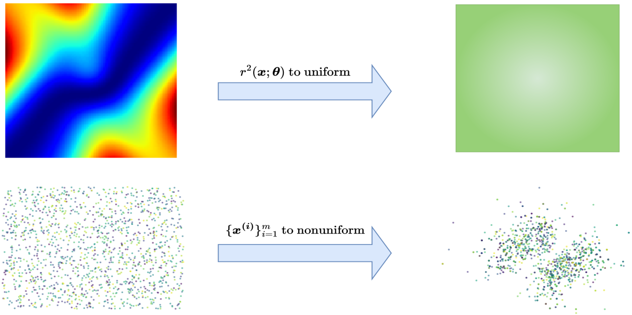
Two things

- minimize the residual:
- endeavor to maintain a smooth profile of the residual

Adversarial adaptive sampling

Two things need to handle together

- minimize the residual: $\min_\theta r(\mathbf{x};\theta)$
- how to maintain a smooth profile of the residual?



Adversarial adaptive sampling

A min-max formulation

- minimize the residual: $\min_\theta r(\mathbf{x};\theta)$
- maintain a smooth profile of the residual

$$\min_{\theta} \max_{p_\alpha \in V} \mathcal{J}(u_\theta, p_\alpha) = \int_\Omega r^2(\mathbf{x};\theta) p_\alpha(\mathbf{x}) d\mathbf{x},$$

For simplicity, we remove the boundary residual term.

$$\min_{\theta} \max_{p \in V} \mathcal{J}(u_\theta, p) = \int_\Omega r^2(\mathbf{x};\theta) p(\mathbf{x}) d\mathbf{x}.$$

where

$$p_\alpha(\mathbf{x}) = p_{\mathbf{Z}}(f_\alpha(\mathbf{x})) |\nabla_{\mathbf{x}} f_\alpha|.$$

is a flow model.

Adversarial adaptive sampling

How can this min-max formulation achieve our goal?

- Optimal transport theory
- Some constraints for V

Wasserstein distance

$$d_{W^M}(\mu, \nu) = \inf_{\pi \in \Pi(\Omega \times \Omega)} \int_{\Omega \times \Omega} d_M(\mathbf{x}, \mathbf{y})\, d\pi(\mathbf{x}, \mathbf{y}),$$

Typically,

$$V := \{p(\boldsymbol{x}) | \|p\|_{\mathrm{Lip}} \leq 1,\, 0 \leq p(\boldsymbol{x}) \leq M\},$$

where M is a positive number, or

$$\hat{V} = \{p(\boldsymbol{x}) | \|p\|_{\mathrm{Lip}} \leq 1,\, p(\boldsymbol{x}) \geq 0, \int_{\Omega} p(\boldsymbol{x}) d\boldsymbol{x} = 1\}.$$

Adversarial adaptive sampling

The min-max formulation

$$\inf_{u} \sup_{p \in \hat{V}} \mathcal{J}(u, p) = \int_{\Omega} r^2(u(\boldsymbol{x})) p(\boldsymbol{x}) \, d\boldsymbol{x},$$

The constraint for p is important.
Otherwise, the maximization step will yield a delta measure

$$\delta(\boldsymbol{x} - \boldsymbol{x}_0) = \arg\max_{p > 0, \int_{\Omega} p d\boldsymbol{x} = 1} \int_{\Omega} r^2(\boldsymbol{x}; \theta) p(\boldsymbol{x}) d\boldsymbol{x},$$

where $\boldsymbol{x}_0 = \arg\max_{\boldsymbol{x} \in \Omega} r^2(\boldsymbol{x}; \theta)$.

Adversarial adaptive sampling

How this maximization step push the residual-induced distribution to a uniform one?

$$
\begin{aligned}
&\sup_{p\in V}\int_\Omega r^2(\boldsymbol{x};\theta)p(\boldsymbol{x})d\boldsymbol{x}\\
=&\sup_{p\in V}\int_\Omega r^2(\boldsymbol{x};\theta)p(\boldsymbol{x})d\boldsymbol{x}-\int_\Omega r^2(\boldsymbol{x};\theta)d\boldsymbol{x}\int_\Omega p(\boldsymbol{x})d\boldsymbol{x}+\int_\Omega r^2(\boldsymbol{x};\theta)d\boldsymbol{x}\int_\Omega p(\boldsymbol{x})d\boldsymbol{x}\\
\le&\int_\Omega r^2(\boldsymbol{x};\theta)d\boldsymbol{x}\left(\sup_{p\in V}\left[\int_\Omega p(\boldsymbol{x})d\mu_r-\int_\Omega p(\boldsymbol{x})d\mu_u\right]+\sup_{p\in V}\int_\Omega p(\boldsymbol{x})d\boldsymbol{x}\right)\\
\le&(d_{W^M}(\mu_r,\mu_u)+M)\int_\Omega r^2(\boldsymbol{x};\theta)\,d\boldsymbol{x},
\end{aligned}
$$

μ_u is a uniform distribution.

Adversarial adaptive sampling

Main theoretical results

Theorem

Under certain conditions, $\lim_{n\to\infty} \mathcal{J}(u_n, p_n) = 0$*, for some sequence of functions* $\{p_n\}_{n=1}^{\infty}$ *satisfying the constraints defined in the min-max formulation. Meanwhile, this optimization sequence has the following two properties:*

1. *The residual sequence* $\{r(u_n)\}_{n=1}^{\infty}$ *of* $\{u_n\}_{n=1}^{\infty}$ *converges to* 0 *in* $L^2(d\mu)$*.*
2. *The renormalized squared residual distributions*

$$d\nu_n \triangleq \frac{r^2(u_n)}{\int_\Omega r^2(u_n(\boldsymbol{x}))\,d\boldsymbol{x}}\,d\mu(\boldsymbol{x})$$

converge to the uniform distribution μ *in the Wasserstein distance* d_{W^M}*.*

Adversarial adaptive sampling

How can we implement the min-max optimization problem?

- the minimization step is straightforward
- the maximization step is not trival because of the constraints

A formulation for practical implementation

$$\min_{\theta} \max_{\substack{p_\alpha > 0, \\ \int_\Omega p_\alpha(\boldsymbol{x}) d\boldsymbol{x} = 1}} \mathcal{J}(u_\theta, p_\alpha) = \int_\Omega r^2(\boldsymbol{x}; \theta) p_\alpha(\boldsymbol{x}) d\boldsymbol{x} - \beta \int_\Omega |\nabla_{\mathbf{x}} p_\alpha(\mathbf{x})|^2 d\boldsymbol{x},$$

This formulation makes that p is well-posed

$$\begin{cases} 2\beta \nabla^2 p^* + r^2(\boldsymbol{x}; \theta) - \frac{1}{|\Omega|} \int_\Omega r^2(\boldsymbol{x}; \theta) d\boldsymbol{x} = 0, & \boldsymbol{x} \in \Omega, \\ \frac{\partial p^*}{\partial \mathbf{n}} = 0, & \boldsymbol{x} \in \partial\Omega. \end{cases}$$

Adversarial adaptive sampling

- minimize the residual

$$\int_{\Omega} r^2\left[u_\theta(\boldsymbol{x})\right] p_\alpha(\boldsymbol{x}) d\boldsymbol{x} \approx \frac{1}{m}\sum_{i=1}^{m} r^2\left[u_\theta(\boldsymbol{x}_\alpha^{(i)})\right]$$

- maximization step

$$\mathcal{J}(u_\theta, p_\alpha) \approx \frac{1}{m}\sum_{i=1}^{m} \frac{r^2\left[u_\theta(\boldsymbol{x}_{\alpha'}^{(i)})\right] p_\alpha(\boldsymbol{x}_{\alpha'}^{(i)})}{p_{\alpha'}(\boldsymbol{x}_{\alpha'}^{(i)})} - \beta \cdot \frac{1}{m}\sum_{i=1}^{m} \frac{|\nabla_{\mathbf{x}} p_\alpha(\boldsymbol{x}_{\alpha'}^{(i)})|^2}{p_{\alpha'}(\boldsymbol{x}_{\alpha'}^{(i)})}$$

Training style is similar to WGAN

- simultaneously optimize the approximate solution and the random samples

Numerical experiments

$$
\begin{aligned}
-\Delta u(\boldsymbol{x}) &= s(\boldsymbol{x}) \quad \text{in } \Omega, \\
u(\boldsymbol{x}) &= g(\boldsymbol{x}) \quad \text{on } \partial\Omega,
\end{aligned}
$$

The reference solution is given by

$$
u(x_1, x_2) = \exp\left(-1000[(x_1 - 0.5)^2 + (x_2 - 0.5)^2]\right),
$$

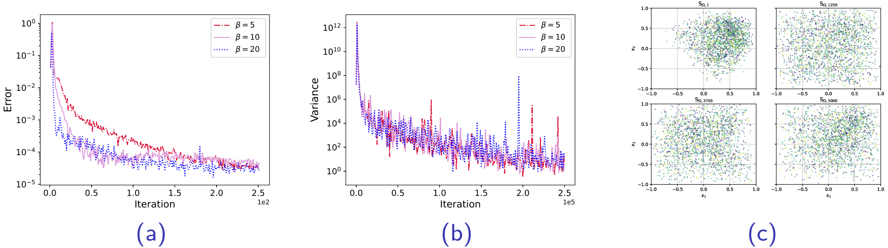
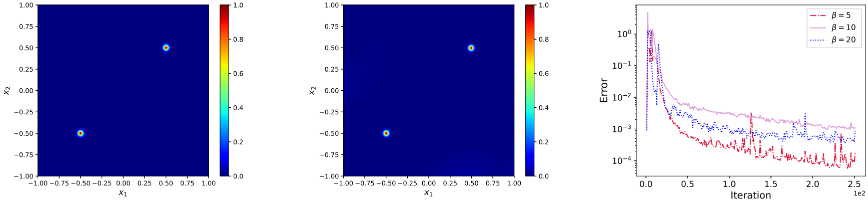


Figure: The results for the peak test problem. (a) The error behaviour. (b) The variance behavior. (c) The evolution of the training set.

Numerical experiments

Two-peak problem

$$
\begin{aligned}
-\nabla \cdot [u(\boldsymbol{x})\nabla v(\boldsymbol{x})] + \nabla^2 u(\boldsymbol{x}) &= s(\boldsymbol{x}) \quad \text{in } \Omega, \\
u(\boldsymbol{x}) &= g(\boldsymbol{x}) \quad \text{on } \partial\Omega,
\end{aligned}
$$



(a) (b) (c)

Figure: The results for the two-peak test problem. (a) The exact solution. (b) AAS approximation. (c) The error behavior.

Numerical experiments

Two-peak problem

$$-\nabla \cdot [u(\boldsymbol{x})\nabla v(\boldsymbol{x})] + \nabla^2 u(\boldsymbol{x}) = s(\boldsymbol{x}) \quad \text{in } \Omega,$$
$$u(\boldsymbol{x}) = g(\boldsymbol{x}) \quad \text{on } \partial\Omega,$$

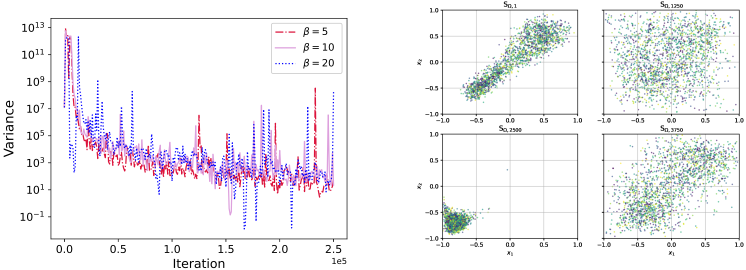
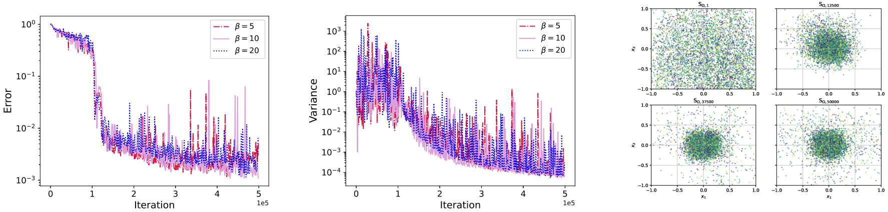


Figure: The evolution of the residual variance and the training set for the two-peak test problem. Left: The variance behavior. Right: The evolution of the training set.

Numerical experiments

$$-\Delta u(\mathbf{x}) + u(\mathbf{x}) - u^3(\mathbf{x}) = s(\mathbf{x}), \quad \mathbf{x} \text{ in } \Omega = [-1,1]^{10}$$
$$u(\boldsymbol{x}) = g(\boldsymbol{x}), \quad \boldsymbol{x} \text{ on } \partial\Omega.$$



(a) (b) (c)

Figure: The results of the ten-dimensional nonlinear test problem. (a) The error behavior. (b) The variance behaviour. (c) The evolution of the training set, $x_1 - x_2$ plane ($\beta = 10$).

Numerical experiments

Table: Error comparison of adaptive sampling methods

Method \ Test problem	One peak	Two peak	High dimensional
PINN	9.74e-04	3.22e-02	1.01
RAR [Lu et. al, 2021]	-	-	9.83e-01
DAS-G	3.75e-04	1.51e-03	9.55e-03
DAS-R	1.93e-04	6.21e-03	1.26e-02
AAS (this work)	**2.97e-05**	**1.09e-04**	**1.31e-03**

Summary of AAS

summary

- the evolution of the training set can be investigated in terms of the optimal transport theory
- a very general and flexible framework for the adaptive learning strategy

outlook

- more robust sampling strategies
- realistic applications

Thank you for your attention
Questions?