Introduction to Physics-Informed Neural Networks and its Convergence

Yoonseo Choi

Uncertainty Quantification Lab
Seoul National University

*Based on: M. Raissi, P. Perdikaris, G.E. Karniadakis, **Physics-informed neural networks**, Journal of Computational Physics 378 (2019) 686–707.*
*Y. Shin, Z. Zhang, G.E. Karniadakis, **Error Estimates of Residual Minimization Using Neural Networks for Linear PDEs**, Journal of Machine Learning for Modeling and Computing. (2023).*

Outline

1. What is a PINN?

2. Error estimate of PINN
    - Continuous RM
    - Discrete RM — Aproach I: Discrete Norm Relation
    - Discrete RM — Aproach II: Rademacher Complexity

3. Summary

What is a PINN?

1. What is a PINN?

A **Physics-Informed Neural Network (PINN)** is a neural network that approximates the solution $u(\mathrm{x}, t)$ of a PDE by incorporating the PDE residual directly into the loss function via **automatic differentiation**.

What is a PDE?

A **Partial Differential Equation (PDE)** is an equation
which involves a **multivariable** function and one or more of its partial derivatives.
Usually initial condition and boundary condition are given.

Examples

- $\frac{\partial f(x,t)}{\partial t} + f(x,t) = 0$
- **Burgers:**

$$\begin{cases} u_t + uu_x - \dfrac{0.01}{\pi} u_{xx} = 0, \quad x \in [-1,1],\ t \in [0,1] \\ \text{boundary condition: } u(0,x) = -\sin(\pi x) \\ \text{initial condition: } u(t,-1) = u(t,1) = 0 \end{cases} \tag{1}$$

- notation: $u_t = \frac{\partial u}{\partial t}$

1. What is a PINN? — Why we need PINN?

Limitations of Classical Numerical Methods

- Traditional methods (FEM, FDM, FVM) require **mesh generation**, which is computationally expensive and difficult in complex geometries.
- Solving **inverse problems** (estimating unknown parameters λ in $\mathcal{N}[u; \lambda] = 0$ from data) is fundamentally different from forward problems and requires separate algorithms.
- High-dimensional PDEs suffer from the **curse of dimensionality**.

So: What is PINN

A **Physics-Informed Neural Network (PINN)** is a neural network that approximates the solution $u(\mathrm{x}, t)$ of a PDE by incorporating the PDE residual directly into the loss function via **automatic differentiation**.

- **Meshfree**: no mesh required; collocation points can be sampled freely.
- **Exact derivatives**: auto-diff computes $\partial u/\partial x$, $\partial^2 u/\partial x^2$, $\ldots$ exactly (not finite-difference approximations), using the same backpropagation graph.
- **Unified forward/inverse framework**: unknown parameters λ are simply added as trainable variables alongside the network weights θ.

1. What is a PINN? — Process

Process of PINN

Problem Setting

Given $\mathcal{N}[u(x,t);\lambda] = 0$ (with IC, BC), the goal is to find $u(x,t)$. ($\mathcal{N}$: differential operator)

Process

Approximate the solution $u(\mathrm{x},t)$ by a deep neural network using the PINN loss function:

$$\mathcal{L}(\theta) = \underbrace{\mathcal{L}_{\mathrm{data}}}_{\text{boundary/initial}} + \underbrace{\mathcal{L}_f}_{\text{PDE residual}}$$

$$\mathcal{L}_{\mathrm{data}} = \frac{1}{N_u}\sum_{i=1}^{N_u}\left|\hat{u}(\mathrm{point}_i) - u^i\right|^2, \quad \mathrm{point}_i \in \text{initial/boundary domain}$$

$$\mathcal{L}_f = \frac{1}{N_f}\sum_{i=1}^{N_f}\left|\mathcal{N}(u(\mathrm{point}_i))\right|^2, \quad \mathrm{point}_i \in \text{collocation domain}$$

1. What is a PINN? — Does it work?

Intuition

$\mathcal{L}_{\text{data}}$ is the standard data-fitting loss; $\mathcal{L}_f$ acts as a **regularization** that penalizes solutions violating the governing equation at collocation points.

Does it work?

Yes — empirically PINNs approximate PDE solutions accurately. But. . .

Open questions:

- Does the neural network $u_{\hat{\theta}}$ truly converge to the real solution u^*?
- How accurate is it, and can we *quantify* the error?
- What happens as the network grows larger and more collocation points are used?

Goal of Section 2

We study the paper: *Shin, Zhang & Karniadakis (2023)* which provides a rigorous **abstract framework** for analyzing the convergence and error estimates of residual minimization using neural networks for **linear PDEs**.

Error estimate of PINN

2. Error estimate of PINN

Main Question

- 신경망(NN)이 주어진 데이터에 대해서는 완벽하게 학습되었다면, NN과 PDE의 진짜 해 간 차이는 어느정도인가? (posterior, priori estimate)
- 0으로 수렴하는가? (Convergence)

Outline

- 2.1 continuous residual minimization
- 2.2 discrete residual minimization
    - Bernstein type inequality가 성립하는 NN
    - 일반적인 NN

2. Error estimate of PINN—Mathematical Setup

Problem Setting

Let $A : X \to Y$ and $B : X \to Z$ be linear operators on Banach spaces $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$, $(Z, \|\cdot\|_Z)$. Consider the linear problem:

$$Au(x) = f(x),\ x \in \Omega; \qquad Bu(x) = g(x),\ x \in \Gamma,$$

where $\Omega \subset \mathbb{R}^d$ is a bounded domain and $\Gamma = \partial\Omega$.

Definition 1.1 Solution of PDE

Let $(V, \|\cdot\|_V)$ be a Banach space with $(X, \|\cdot\|_X)$ a dense subspace. $u^* \in V$ is a **solution** to the problem if there exists a sequence $\{u_k^*\} \subset X$ such that

$$\lim_{k\to\infty} \|u_k^* - u^*\|_V = 0 \quad \text{and} \quad \lim_{k\to\infty} \|Au_k^* - f\|_Y + \|Bu_k^* - g\|_Z = 0.$$

2. Error estimate of PINN—Mathematical Setup

Model Setting

Focus on feed forward Neural Network. Notation of NN and NN class is as follow:

- $\mathcal{N}_{\theta,n}$: class of NN whose size is less than n.
- $u_{N,n}$: one NN whose size is less than n.

More precisely,

- $\mathcal{R}[\theta]$: L-layer feed forward Neural Network from $\mathbb{R}^{n_0}$ to $\mathbb{R}^{n_L}$
- $\vec{n}$: number of Neurons of NN $(= (n_0, \dots, n_L))$
- $$\Theta(\vec{n}) = \left\{ \{(W_j, b_j)\}_{j=1}^{L} : W_j \in \mathbb{R}^{n_j \times n_{j-1}},\ b_j \in \mathbb{R}^{n_j},\ j = 1, 2, \dots, L \right\}.$$
  : the set of all possible network parameters with the fixed architecture $\vec{n}$
- $$\mathcal{N}_{\theta,n} = \{\mathcal{R}[\theta](x) : \theta \in \cup_{\vec{v} \le \vec{n}_n} \Theta(\vec{v})\}.$$
  : $\vec{n}_n$ is the size of NN in stage n. Gets bigger($\vec{n}_n < \vec{n}_{n+1}$
  : By construction, $\mathcal{N}_{\theta,n} \subset \mathcal{N}_{\theta,n+1}$.
- $u_{N,n}$: An element of $\mathcal{N}_{\theta,n}$

Error estimate of PINN

Continuous RM

2. 1 Continuous RM — Setting

Loss Function Setting

Consider the **continuous** loss functional ($\tau > 0$, $p \geq 1$):

$$J_\tau(v) = \|f - Av\|_Y^p + \tau\|g - Bv\|_Z^p, \qquad \inf_{v \in \mathcal{N}_{\theta,n} \cap X} J_\tau(v).$$

2. 1 Continuous RM — Assumptions

Assumption 1.1 (Existence)

Ass. 1.1: There exists a solution $u^* \in V$ to the problem in the sense of Def. 1.1

Assumption 1.2 (Norm Relations)

Ass. 1.2: The operators A, B satisfy $\|Av\|_Y, \|Bv\|_Z < \infty$ for all $v \in V$, and:

$$C_1\|u\|_V \leq \|Au\|_Y + \|Bu\|_Z, \quad \forall u \in X \qquad \text{(2a: stability of solution)}$$

$$\|Au\|_Y + \|Bu\|_Z \leq C_2\|u\|_X, \quad \forall u \in X \qquad \text{(2b: smoothness of solution)}$$

where $C_1, C_2 > 0$ are independent of u.

Under ass. 1.1 and 1.2, solution $u*$ is unique.

Assumption 1.3. (Universal NN Approximation)

There exists a sequence of NN classes $\mathcal{N}_{\theta,n} \subset \mathcal{N}_{\theta,n+1}$ such that $X \subset \overline{\bigcup_n \mathcal{N}_{\theta,n}}$ in the topology of $(X, \|\cdot\|_X)$.

2. 1 Continuous RM — Assumptions

Are the assumptions unreal?

Assumption 1.2 (Norm Relations)

Ass. 1.2: Typical assumption. Holds true for Well-posedness PDE or linear 2nd PDEs.

Assumption 1.3. (Universal NN Approximation)

Make sense. Similar to Universal Approximation Theorem

2. 1 Continuous RM — Error Estimates (Thm. 1.1)

Theorem 1.1 (최선의 NN의 오차 상한)

Let Ass. 1.1(existence), 1.2(norm relation) hold. Let $u^{\tau}_{N,n} \in \mathcal{N}_{\theta,n} \cap X$ minimize J_τ and u^* be the solution. For $\tau \geq 1$:

A posterior estimate:

$$\|u^{\tau}_{N,n} - u^*\|_V \;\leq\; C_1^{-1} \cdot 2^{\frac{p-1}{p}} \left(J_\tau(u^{\tau}_{N,n})\right)^{1/p}.$$

A priori estimate: For any $\varepsilon > 0$, $\exists\, u^*_\varepsilon \in X$ such that

$$\|u^{\tau}_{N,n} - u^*\|_V \;\leq\; \frac{2^{\frac{p-1}{p}}(1+\tau)^{1/p}}{C_1}\left(C_2 \inf_{w \in \mathcal{N}_{\theta,n} \cap X} \|w - u^*_\varepsilon\|_X + \varepsilon\right).$$

If $u^* \in X$, then $u^*_\epsilon = u^*$.

참고

posterior estimate: 신경망을 학습하지 않고 신경망 구조만을 이용해 얻을 수 있는 오차 한계
priori estimate: 신경망 학습으로 구한 최적의 NN으로 계산하는 오차 한계

2. 1 Continuous RM — Error Estimates (Thm. 1.1)

Proof of Theorem 1.1(posterior estimate)

Step 1. **(Use ass. 1.2)** For $\tau \geq 1$ and any $w \in X$, by
(2a):$C_1\|u^* - w\|_V \leq \|A(u^* - w)\|_Y + \|B(u^* - w)\|_Z = \|f - Aw\|_Y + \|g - Bw\|_Z$

Step 2. **Use** $(a+b)^p \leq 2^{p-1}(a^p + b^p)$ Then
$(\|f - Aw\|_Y + \|g - Bw\|_Z)^p \leq 2^{p-1}(\|A(u^* - w)\|_Y^p + \|g - Bw\|_Z^p) = 2^{p-1}(J_1(w))^{1/p} \leq 2^{(p-1)} J_\tau(w)$.
Setting $w = u^\tau_{N,n}$ and taking both sides to the 1/p power yields the a posteriori bound.

Proof of Theorem 1.1(priori estimate)

Step 1. **(A priori, bounding J_τ)**

By posterior estimate, $C_1\|u^\tau_{N,n} - u^*\|_V \leq C_1^{-1} \cdot 2^{\frac{p-1}{p}} \left(J_\tau(u^\tau_{N,n})\right)^{1/p}$.
Since $u^\tau_{N,n}$ minimizes over $\mathcal{N}_{\theta,n} \cap X$, $J_\tau(u^\tau_{N,n}) \leq \inf{}_w J_\tau(w)$.
Also, $A^p + \tau B^p \leq (1+b)^p + \tau(A+B)^p = (1+\tau)(A+B)^p$ Combining all,
$C_1\|u^\tau_{N,n} - u^*\|_V \leq 2^{(p-1)/p}(1+\tau)^{1/p} \inf_w \big(\|Aw - f\|_Y + \|Bw - g\|_Z\big)$.

Step 2. **(Triangle inequality + (2b))**
Let $u^*_\varepsilon \in X$ satisfy $\|Au^*_\varepsilon - f\|_Y + \|Bu^*_\varepsilon - g\|_Z < \varepsilon$ (exists by Ass. 1.1).
Then $\|Aw - f\|_Y + \|Bw - g\|_Z \leq C_2\|w - u^*_\varepsilon\|_X + \varepsilon$ by triangle inequality and (2b).
Combining with Step 1 gives the a priori bound. □

2. 1 Continuous RM — Convergence (Thm. 1.2)

Theorem 1.2 (최선의 NN와 해 간의 차이가 0으로 수렴한다.)

Suppose Ass. 1.1, 1.2, and 1.3 hold. For fixed $\tau \geq 1$, let $u^{\tau}_{N,n}$ be a quasi-minimizer of J_{τ}, i.e $J_{\tau}(u^{\tau}_{N,n}) \leq \inf_{v \in \mathcal{N}_{\theta,n} \cap X} J_{\tau}(v) + \delta_n$ with $\delta_n \to 0$. Then:

$$\lim_{n \to \infty} \|u^{\tau}_{N,n} - u^*\|_V = 0.$$

Proof of Theorem 1.2

Combine Proposition 1.3 and Theorem 1.1 directly:

Step 1. By the **a posteriori estimate** of Thm. 1.1:

$$\|u^{\tau}_{N,n} - u^*\|_V \leq C_1^{-1} \cdot 2^{(p-1)/p} \big(J_{\tau}(u^{\tau}_{N,n})\big)^{1/p}.$$

Step 2. By Prop. 1.3:

$$\lim_{n \to \infty} J_{\tau}(u^{\tau}_{N,n}) = 0.$$

Step 3. Combining Steps 1–2:

$$\lim_{n \to \infty} \|u^{\tau}_{N,n} - u^*\|_V = 0. \quad \square$$

2. 1 Continuous RM — Convergence (Thm. 1.2)

Proposition 1.3 (신경망의 크기가 커지면 continuous 버전의 loss는 0으로 수렴한다)

For fixed $\tau \geq 1$, let $u^{\tau}_{N,n}$ be a quasi-minimizer of J_τUnder Ass. 1.1, 1.3, and (2b) of Ass. 1.2, $\lim_{n\to\infty} J_\tau(u^{\tau}_{N,n}) = 0$.

Proof of Proposition 1.3

Step 1 **Setting**

Let v be the exact solution to the PDE and $\{v_k^*\}$ be corresponding sequence in X. (def 1.1)
By Ass. 1.3, for a sufficiently large k, $\exists n_k$ and a $\exists$ corresponding $u_{n_k} \in \mathfrak{N}_{\theta,n_k}$ such that:

$$||v_k^* - u_{n_k}||_X \leq \epsilon_k \ (\epsilon_k : \text{positive sequence converging to } 0)$$

Step 2 $\lim_{k\to\infty} \mathcal{J}_\tau(u^{\tau}_{\mathbb{N},n_k}) = 0$

Using **Ass. 1.2 (2b)**,

$$\mathcal{J}_\tau(u_{n_k}) \leq (||f - Av_k^*||_Y + C_2\epsilon_k)^p + \tau(||g - Bv_k^*||_Z + C_2\epsilon_k)^p$$

By def of quasi-minimizer,

$$\mathcal{J}_\tau(u^{\tau}_{\mathbb{N},n_k}) \leq \mathcal{J}_\tau(u_{n_k}) + \delta_{n_k}$$

Taking $k \to \infty$, the right-hand side converges to 0, yielding $\lim_{k\to\infty} \mathcal{J}_\tau(u^{\tau}_{\mathbb{N},n_k}) = 0$.

2. 1 Continuous RM — Convergence (Thm. 1.2)

Proof of Proposition 1.3

Step 3: Monotonic Upper-Bound via Space Inclusion

For any network sizes $m_1 < m_2$, since $\mathfrak{N}_{\theta,m_{\mathbf{1}}} \subset \mathfrak{N}_{\theta,m_{\mathbf{2}}}$,

$$\mathcal{J}_\tau(u^\tau_{\mathbb{N},m_{\mathbf{2}}}) \leq \mathcal{J}_\tau(u^\tau_{\mathbb{N},m_{\mathbf{1}}}) + \delta_{m_{\mathbf{2}}}$$

For any $\epsilon > 0$, we choose a n_K such that $\mathcal{J}_\tau(u^\tau_{\mathbb{N},n_K}) \leq \epsilon/2$.
For any arbitrary $n \geq n_K$,

$$\mathcal{J}_\tau(u^\tau_{\mathbb{N},n}) \leq \mathcal{J}_\tau(u^\tau_{\mathbb{N},n_K}) + \delta_n \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Since this holds for all $n \geq n_K$, we conclude that the entire sequence converges to 0:

$$\lim_{n\to\infty} \mathcal{J}_\tau(u^\tau_{\mathbb{N},n}) = 0$$

Error estimate of PINN

Discrete RM — Aproach I: Discrete Norm Relation

2. 2 Discrete RM — Setting

Definition: Discrete Function Spaces and Norms

Let $Y = L^p(\Omega)$, $Z = L^p(\Gamma)$, and fix training points $\{(x_i^r, w_i^r)\}_{i=1}^{M_r} \subset \Omega \times \mathbb{R}_{>0}$ and $\{(x_i^b, w_i^b)\}_{i=1}^{M_b} \subset \Gamma \times \mathbb{R}_{>0}$.

Discrete Norm $\|v\|_{Y_{M_r}} := \left(\sum_{i=1}^{M_r} w_i^r \, v(x_i^r)^p\right)^{1/p}$
Discrete Function Space $Y_{M_r} := \{v \in Y : v(x_i^r)$ is continuous at each x_i^r and $\|v\|_{Y_{M_r}} < \infty.\}$

Similarly, Z_{M_b} with norm $\|v\|_{Z_{M_b}} := (\sum_{i=1}^{M_b} w_i^b \, v(x_i^b)^p)^{1/p}$.

Loss Function Setting

Let $Y = L^p(\Omega)$, $Z = L^p(\Gamma)$. The discretized loss is (with $M = (M_r, M_b)$, $\tau \geq 1$):

$$J_\tau^{\mathrm{M}}(v) = \sum_{i=1}^{M_r} w_i^r \left(f(x_i^r) - Av(x_i^r)\right)^p + \tau \sum_{i=1}^{M_b} w_i^b \left(g(x_i^b) - Bv(x_i^b)\right)^p.$$

Using the notation above, the discrete loss functional (12) is written compactly as

$$J_\tau^M(v) = \|f - Av\|_{Y_{M_r}}^p + \tau \, \|g - Bv\|_{Z_{M_b}}^p.$$

When $p = 2$, $w_i^r = 1/M_r$, $w_i^b = 1/M_b$: this recovers the standard PINN loss.

2. 2 Discrete RM — Approach I: Discrete Norm Relation

Assumption 2.1 (Bernstein-type Discrete Norm Relation)

For any $n \in \mathbb{N}$, there exist M_r, M_b (depending on n) such that

$$\|Av\|_{Y_{M_r}} \geq \frac{1}{2}\|Av\|_Y, \quad \forall v \in \{v \in \mathcal{N}_{\theta,n} \cap X : Av \in Y_{M_r}\},$$

$$\|Bv\|_{Z_{M_b}} \geq \frac{1}{2}\|Bv\|_Z, \quad \forall v \in \{v \in \mathcal{N}_{\theta,n} \cap X : Bv \in Z_{M_b}\}.$$

This is a **Bernstein-type inequality**: the discrete norm controls the continuous norm on the NN class. It holds for **Gaussian RBF networks** (see Example 4.2) but is *unavailable* for general deep ReLU networks.

Theorem 2.1 (최선의 NN과 해 간 차이의 오차 상한)

Let $Y = L^2(\Omega)$, $Z = L^2(\Gamma)$, $V = X$. Let Ass. 1.2, 2.1 hold. Let $\|v\|_{Y_{M_r}} \leq C_3\|v\|_Y$ and $\|v\|_{Z_{M_b}} \leq C_3\|v\|_Z$ with C_3 independent of M_r, M_b. Let $u_{N,n}^{\tau,M}$ minimize J_τ^M over $\mathcal{N}_{\theta,n} \cap X$. Then:

$$\|u_{N,n}^{\tau,M} - u^*\|_V \leq 2\sqrt{2}\, C_1^{-1} \big(J_\tau^M(u_{N,n}^{\tau,M})\big)^{1/2} + 3\sqrt{2}\, C_1^{-1} C_3\, \varepsilon_{f,g,u}^{1/2},$$

where $\varepsilon_{f,g,u} = \inf_{\tilde{v}}(\|A\tilde{v} - f\|_Y^2 + \|B\tilde{v} - g\|_Z^2)$, $\tilde{v} - u_{N,n}^{\tau,M} \in \tilde{V}_{n_1}$.
$\tilde{V}_{n_1} := \{v \in \mathcal{N}_{\theta,n_1} \cap X : Av \in Y_{M_r},\ Bv \in Z_{M_b}\}$.

2. 2 Discrete RM — Approach I: Discrete Norm Relation

Proof of Theorem 2.1

Let $\tilde{v} \in X$ with $\tilde{v} - u_{N,n}^{\tau,M} \in \tilde{V}_{n_1}$.

Step 1. (Apply Ass. 2.3 and 4.1 to $\tilde{v} - u_{N,n}^{\tau,M}$)
By (2a) and Ass. 4.1 ($\|\cdot\|_{Y_{M_r}} \geq \frac{1}{2}\|\cdot\|_Y$ on $\tilde{V}_{n_1}$):

$$\begin{aligned} C_1\|\tilde{v} - u_{N,n}^{\tau,M}\|_V &\leq \|A(\tilde{v} - u_{N,n}^{\tau,M})\|_Y + \|B(\tilde{v} - u_{N,n}^{\tau,M})\|_Z \\ &\leq 2\|A(\tilde{v} - u_{N,n}^{\tau,M})\|_{Y_{M_r}} + 2\|B(\tilde{v} - u_{N,n}^{\tau,M})\|_{Z_{M_b}}. \end{aligned}$$

Step 2. (Triangle inequality on discrete norms)
$\|A(\tilde{v} - u_{N,n}^{\tau,M})\|_{Y_{M_r}} \leq \|f - Au_{N,n}^{\tau,M}\|_{Y_{M_r}} + (J_1^M(\tilde{v}))^{1/2}$ (Cauchy–Schwarz).
Thus: $C_1\|\tilde{v} - u_{N,n}^{\tau,M}\|_V \leq 2\sqrt{2}(J_1^M(u_{N,n}^{\tau,M}))^{1/2} + 2\sqrt{2}(J_1^M(\tilde{v}))^{1/2}$.

Step 3. (Triangle inequality for u^*)
$\|u^* - u_{N,n}^{\tau,M}\|_V \leq \|u^* - \tilde{v}\|_V + \|\tilde{v} - u_{N,n}^{\tau,M}\|_V$.
Apply Ass. 2.3 to bound $\|u^* - \tilde{v}\|_V \leq C_1^{-1}(J_1^M(\tilde{v}))^{1/2}$ (using $\|v\|_{Y_{M_r}} \leq C_3\|v\|_Y$).

Step 4. (Combine) Taking infimum over valid $\tilde{v}$ yields the stated bound. □

2. 2 Discrete RM — Approach I: Discrete Norm Relation

- 확률적으로 상한이 존재하는 것이 아니므로 꽤 강력한 증명
- 하지만 ass. 2.1이 현실적으로 강한 가정.
- 따라서, Rademacher Complexity를 이용해 좀 더 완화된 가정 하에서도 성립하는 오차 상한 구함.

Error estimate of PINN

Discrete RM — Aproach II: Rademacher Complexity

2. 3 Discrete RM — Approach II: Rademacher Complexity

Definition (Rademacher Complexity)

Given i.i.d. samples $\{X_i\}_{i=1}^{M}$, the **Rademacher complexity** of a function class $\mathcal{F}$ is:

$$R_M(\mathcal{F}) = \mathbb{E}_{\{X_i,\varepsilon_i\}}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{M}\sum_{i=1}^{M}\varepsilon_i f(X_i)\right|\right],$$

where ε_i are i.i.d. Rademacher RVs: $\mathbb{P}(\varepsilon_i = \pm 1) = 0.5$.
For two-layer tanh networks with bounded weights: $R_M(\mathcal{N}) \leq \frac{\omega_{\max}}{\sqrt{M}}(1 + 2\gamma\omega_{\max}(1 + \sqrt{2\log(2d)}))$.

- 모델이 random noise를 다 외울만큼 유연한가?를 나타내는 지표 (0에 가까울수록 모델의 유연성이 떨어짐.)
- ϵ이 1일 때 $f(X_i)$가 양수이고, ϵ이 -1일 때 $f(X_i)$가 음수라면, 즉 모델이 랜덤하게 1 또는 -1을 출력하는 ϵ과 패턴이 같다면 rademacher complexity가 커짐.
- 주로 Rademacher Complexity의 상한을 구함. (신경망의 크기/$\sqrt{pt0ø,}$꼴)

2. 3 Discrete RM — Approach II: Rademacher Complexity

Assumption 2.2

Let $G_r \geq \max\{\|f\|_{L^\infty(\Omega)}, \sup_k \|Au_k^* - f\|_{L^\infty(\Omega)}\}$ and $G_b \geq \max\{\|g\|_{L^\infty}, \sup_k \|Bu_k^* - g\|_{L^\infty}\}$. Assume $G_r, G_b < \infty$.

(* 현실적인 가정)

New Function Classes

Define the **function classes** (using rational-weight NN subclass $\mathcal{N}_{\theta,n}^{Q}$):

$$\mathcal{F}_{r,n} := \{Av - f \in Y : v \in \mathcal{N}_{\theta,n}^{Q} \cap X,\ \|Av - f\|_{L^\infty(\Omega)} \leq G_r\},$$

$$\mathcal{F}_{b,n} := \{Bv - g \in Z : v \in \mathcal{N}_{\theta,n}^{Q} \cap X,\ \|Bv - g\|_{L^\infty(\Gamma)} \leq G_b\},$$

$$\widetilde{\mathcal{N}}_{\theta,n}^{Q} := \{v \in \mathcal{N}_{\theta,n}^{Q} \cap X : Av - f \in \mathcal{F}_{r,n},\ Bv - g \in \mathcal{F}_{b,n}\}.$$

즉, 해와의 오차가 특정 값 이하인, '다루기 수월한' NN만 모아둔 subclass.
Approach II는 이 subclass에 포함되는 NN에 대해 전개된다.

2. 3 Discrete RM (Approach II) — Lemma 2.2

Lemma 2.2 (discrete loss와 continuous loss간 차이의 상한은 Rademacher complexity로 표현할 수 있다.)

Suppose Ass. 2.2 holds. Let $\{x_i^r\}_{i=1}^{M_r}$, $\{x_i^b\}_{i=1}^{M_b}$ be i.i.d. with densities ρ over Ω and ρ_b over Γ. Let $Y = L^p_\rho(\Omega)$, $Z = L^p_{\rho_b}(\Gamma)$. For any $\delta_r, \delta_b > 0$, **with probabilities** Q_r, Q_b respectively:

$$\sup_{Av-f\in\mathcal{F}_{r,n}} \Big| \|Av - f\|^p_{Y_{M_r}} - \|Av - f\|^p_{L^p_\rho(\Omega)} \Big| \le 2R_{M_r}(\mathcal{F}^p_{r,n}) + \frac{\delta_r}{2},$$

$$\sup_{Bv-g\in\mathcal{F}_{b,n}} \Big| \|Bv - g\|^p_{Z_{M_b}} - \|Bv - g\|^p_{L^p_{\rho_b}(\Gamma)} \Big| \le 2R_{M_b}(\mathcal{F}^p_{b,n}) + \frac{\delta_b}{2},$$

where $Q_r = 1 - 2\exp\!\big(-\frac{M_r\delta_r^2}{32G_r^{2p}}\big)$, $Q_b = 1 - 2\exp\!\big(-\frac{M_b\delta_b^2}{32G_b^{2p}}\big)$.

Combining both,

$$\sup_{v\in\widetilde{\mathcal{N}}^Q_{\theta,n}} \big| J^M_\tau(v) - J_\tau(v) \big| \le 2R_{M_r}(\mathcal{F}^p_{r,n}) + 2\tau R_{M_b}(\mathcal{F}^p_{b,n}) + \frac{\delta_r}{2} + \tau\frac{\delta_b}{2} withwithprob. \ \ge Q_r Q_b$$

intuition rademacher complexity가 잘 통제된다면 discrete loss와 continuous loss가 가까워진다.

2. 3 Discrete RM (Approach II) — Proof of Lemma 2.2

Proof of Lemma 2.2

Step 1. **(Rewrite as empirical mean)**
Observe that $\|Av - f\|_{Y_{M_r}}^p = \frac{1}{M_r}\sum_{i=1}^{M_r}(f(x_i^r) - Av(x_i^r))^p$ is the empirical mean of $(f - Av)^p$ at i.i.d. samples drawn from ρ.

Step 2. **(Apply uniform law of large numbers to $\mathcal{F}_{r,n}$)**
Since $\mathcal{F}_{r,n}$ is countable and $P(\sup_{F\in\mathcal{F}_{r,n}}|F(x)| \le G_r) = 1$ (by continuity of probability measure), apply the uniform law of large numbers via Rademacher complexity :

$$\sup_{Av-f\in\mathcal{F}_{r,n}}\left|\frac{1}{M_r}\sum_i (f(x_i^r) - Av(x_i^r))^p - \mathbb{E}[(f - Av)^p]\right| \le 2R_{M_r}(\mathcal{F}_{r,n}^p) + \frac{\delta_r}{2}$$

with probability $Q_r = 1 - 2\exp(-M_r\delta_r^2/(32G_r^{2p}))$.
* NN을 앞에서 정의한 subclass로 제한했기에 가능한 이야기

Step 3. **(Repeat for $\mathcal{F}_{b,n}$ and combine)**
Apply the same argument to $\mathcal{F}_{b,n}$ with bound G_b, obtaining probability Q_b. Combine by the decomposition $|J_\tau^M(v) - J_\tau(v)| \le |\cdots|_r + \tau|\cdots|_b$; a union bound gives joint probability Q_rQ_b. □

2. 3 Discrete RM (Approach II) — Proof of Lemma 2.2

Theorem (uniform law of large numbers via Rademacher complexity)

Suppose $\forall f \in \mathcal{F}$, $\|f\|_\infty \le b$, i.e., the function class is uniformly bounded. Then with probability $1-\delta$,

$$\frac{1}{2}\mathcal{R}_n(\bar{\mathcal{F}}) - b\sqrt{\frac{2\log(2/\delta)}{n}} \;\le\; \|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} \;\le\; 2\mathcal{R}_n(\mathcal{F}) + b\sqrt{\frac{2\log(2/\delta)}{n}},$$

where $\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} := \sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}[f]\right|$ and $\bar{\mathcal{F}} := \{f - f' : f, f' \in \mathcal{F}\}$.

2. 3 Discrete RM (Approach II)— Thm. 2.3

Theorem 2.3 (최선의 NN의 오차 상한)

Suppose Ass. 1.1, 1.2, 2.2 hold. Let $\{x_i^r\}, \{x_i^b\}$ be i.i.d. with densities ρ, ρ_b. Let $\tau \geq 1$, $u_{N,n}^{\tau,M}$ minimize J_τ^M over $\widetilde{\mathcal{N}}_{\theta,n}^Q$, and $\widetilde{R}_M(\mathcal{F}_n^p) := R_{M_r}(\mathcal{F}_{r,n}^p) + \tau R_{M_b}(\mathcal{F}_{b,n}^p)$. **A posterior estimate** For $\delta > 0$, with prob. $\geq Q_{r,b} = \left(1 - 2e^{-\frac{M_r \delta^2}{32 G_r^{2p}}}\right)\left(1 - 2e^{-\frac{M_b \delta^2}{32 G_b^{2p}}}\right)$:

$$\|u_{N,n}^{\tau,M} - u^*\|_V \leq C_1^{-1} 2^{\frac{p-1}{p}} \left[J_\tau^M(u_{N,n}^{\tau,M}) + 2\widetilde{R}_M(\mathcal{F}_n^p) + (1+\tau)\delta/2\right]^{1/p}.$$

A priori estimate For any $\varepsilon > 0$, $\exists\, u_\varepsilon^* \in X$ s.t. with prob. $\geq Q_{r,b}$:

$$\|u_{N,n}^{\tau,M} - u^*\|_V \leq C_1^{-1} 2^{\frac{p-1}{p}} \left[(1+\tau) \inf_{w \in \widetilde{\mathcal{N}}_{\theta,n}^Q} (C_2 \|w - u_\varepsilon^*\|_X + \varepsilon)^p + 4\widetilde{R}_M(\mathcal{F}_n^p) + (1+\tau)\delta\right]^{1/p}.$$

intuition train loss를 작게 하고, rademacher complexity를 통제할 수 있다면 해와 NN 간의 오차를 통제할 수 있다.

2. 3 Discrete RM (Approach II)— Proof of Theorem 2.3

Proof of Theorem 2.3

Easy. Just use former thm and assumptions.

Step 1. **(A posteriori bound from continuous RM)**
By Thm. 3.2 (a posteriori): $\|u_{N,n}^{\tau,M} - u^*\|_V \le C_1^{-1} 2^{(p-1)/p} (J_\tau(u_{N,n}^{\tau,M}))^{1/p}$.

Step 2. **(Bridge continuous and discrete loss via Lemma 4.5)**
Since $u_{N,n}^{\tau,M} \in \widetilde{\mathcal{N}}_{\theta,n}^{Q}$, Lemma 4.5 gives with prob. $Q_{r,b}$:

$$J_\tau(u_{N,n}^{\tau,M}) \le J_\tau^M(u_{N,n}^{\tau,M}) + 2\widetilde{R}_M(\mathcal{F}_n^p) + (1+\tau)\delta/2.$$

Combining with Step 1 yields the first (a posteriori) bound.

Step 3. **(A priori bound)**
By minimality of $u_{N,n}^{\tau,M}$: $J_\tau^M(u_{N,n}^{\tau,M}) \le \inf_{w \in \widetilde{\mathcal{N}}_{\theta,n}^{Q}} J_\tau^M(w)$.

Apply Lemma 4.5 again to bound $\inf_w J_\tau^M(w) \le \inf_w J_\tau(w) + 2\widetilde{R}_M(\mathcal{F}_n^p) + (1+\tau)\delta/2$.
Then by (2b) and Ass. 2.2: $\inf_w J_\tau(w) \le (1+\tau)(C_2\|w - u_\varepsilon^*\|_X + \varepsilon)^p$.
Combining gives the a priori bound. ☐

2. 3 Discrete RM (Approach II)— Convergence (Thm. 2.4)

Theorem 2.4 (데이터 수가 많아지고 신경망 크기가 커지면 NN이 해로 수렴한다.)

Under the same conditions as Thm. 2.3, suppose further that $\lim_{M_r\to\infty} R_{M_r}(\mathcal{F}^p_{r,n}) = 0$ and $\lim_{M_b\to\infty} R_{M_b}(\mathcal{F}^p_{b,n}) = 0$ for all n, and that $\exists$ a solution sequence $\{v_k^*\} \subset \bigcup_{n=1}^{\infty} \widetilde{\mathcal{N}}^Q_{\theta,n}$ in the topology of $(X, \|\cdot\|_X)$. Then:

$$\lim_{n\to\infty} \lim_{M\to\infty} \|u^{\tau,M}_{N,n} - u^*\|_V = 0, \quad M = (M_r, M_b), \quad \text{in probability.}$$

Proof of Theorem 2.4 (Steps 1–2)

Easy. Just use former thm and assumptions.

Step 1. **(NN approximation of solution sequence)**
By assumption, $\exists$ n_k and $u^Q_{n_k} \in \widetilde{\mathcal{N}}^Q_{\theta,n_k}$ with $\|u^Q_{n_k} - v_k^*\|_X \le \varepsilon_k \to 0$.

Step 2. **(Apply Thm. 2.3 with carefully chosen δ)**
Choose $\delta_r = 2M_r^{-1/2+\varepsilon}$, $\delta_b = 2M_b^{-1/2+\varepsilon}$ for $0 < \varepsilon < 1/2$. By Thm. 2.3, with prob. $\ge (1 - 2e^{-M_r^{\varepsilon}/8G_r^{2p}})(1 - 2e^{-M_b^{\varepsilon}/8G_b^{2p}}) \to 1$:

$$\|u^{\tau,M}_{N,n_k} - u^*\|_V \le C_1^{-1} 2^{(p-1)/p} \left[J^M_\tau(u^Q_{n_k}) + 2\widetilde{R}_M(\mathcal{F}^p_n) + M_r^{-1/2+\varepsilon} + \tau M_b^{-1/2+\varepsilon} \right]^{1/p}.$$

2. 3 Discrete RM (Approach II)— Proof of Theorem 2.4 (cont.)

Proof of Theorem 4.4 (Steps 3–4)

Step 3. **(Take $M \to \infty$)**
As $M \to \infty$: by assumption $R_M(\mathcal{F}_n^p) \to 0$, and $M^{-1/2+\varepsilon} \to 0$. Hence with probability 1:

$$\lim_{M\to\infty} \|u_{N,n_k}^{\tau,M} - u^*\|_V \le C_1^{-1} 2^{(p-1)/p} (J_\tau(u_{n_k}^Q))^{1/p}.$$

Step 4. **(Take $k \to \infty$)**
By Step 1 and (2b) of Ass. 2.3:

$$J_\tau(u_{n_k}^Q) \le (\|f - Av_k^*\|_Y + C_2\varepsilon_k)^p + \tau(\|g - Bv_k^*\|_Z + C_2\varepsilon_k)^p \to 0 \quad \text{as } k \to \infty.$$

Since $\widetilde{\mathcal{N}}_{\theta,m_1}^Q \subset \widetilde{\mathcal{N}}_{\theta,m_2}^Q$ for $m_1 < m_2$, the quasi-minimizer sequence also converges: $\lim_{n\to\infty} \lim_{M\to\infty} \|u_{N,n}^{\tau,M} - u^*\|_V = 0$ in probability. □

Summary

3. Summary

What is PINN

- PDE의 해를 신경망으로 근사.
- Loss = data loss + PDE residual loss.
- Mesh-free이며 forward/inverse problem을 통합된 framework로 다룸.

Does PINN really approximate true solution?

Setting	Result	Key Assumption
Continuous RM	$\|u^{\tau}_{N,n} - u^*\|_V \to 0$ as $n \to \infty$	Ass. 1.1, 1.2, 1.3
Discrete RM I	Error bound (deterministic)	Ass. 1.2 + Bernstein ineq.
Discrete RM II	$\|u^{\tau,M}_{N,n} - u^*\|_V \to 0$ (in prob.)	Ass. 1.2 + Rademacher

(Shin, Zhang & Karniadakis 2023)

3. Summary

Key Takeaways

(1) NN 크기가 커지면 Continuous loss를 최소화하는 NN과 해 간의 차이가 0으로 수렴한다.

(2) Bernstein 부등식이 성립하는 NN (e.g. Gaussian RBF)은 Discrete Loss를 최소화하는 NN과 해 간의 차이에 대한 결정론적 오차 상한을 구할 수 있다. (비현실적 가정)

(3) 일반 NN의 경우, Rademacher complexity로 Discrete Loss를 최소화하는 NN과 해 간의 차이에 대한 확률적인 상한을 구할 수 있다.

(4) PDE의 종류와 무관한 **abstract framework**이므로, 논문에 나온 assumption만 만족한다면 각자의 연구에 적용 가능하다.

References

- M. Raissi, P. Perdikaris, G.E. Karniadakis, *Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations*, Journal of Computational Physics, 378 (2019), 686–707.
- Y. Shin, Z. Zhang, G.E. Karniadakis, *Error Estimates of Residual Minimization Using Neural Networks for Linear PDEs*, Journal of Machine Learning for Modeling and Computing (2023).
- S. Mei, *STAT 210B: Theoretical Statistics, Lecture Notes*, University of California, Berkeley, Spring 2023.