Lecture 15: Operator Learning Theory

Haizhao Yang
Department of Mathematics
University of Maryland College Park

2022 Summer Mini Course
Tianyuan Mathematical Center in Central China

Why Operator Learning?

Broad applications

- Reduced order modeling: learning operators in lower dim
- Solving parametric PDEs
- Solving inverse problems
- Density function theory: potential function to density function
- Phase retrieval: data to images
- Image processing: image to image
- Predictive data science: historical states to future states

Probably most mappings are high-dim or even infinite-dim

Learning Mathematical Operators

Notations

- Function spaces $\mathcal{X}$ and $\mathcal{Y}$, e.g., $\mathbb{R}$-valued over domain $\Omega \subset \mathbb{R}^D$
- Operator $\Psi : \mathcal{X} \to \mathcal{Y}$
- Data samples $\mathcal{S} = \{u_i, v_i\}_{i=1}^{2n}$ with

$$v_i = \Psi(u_i) + \epsilon_i,$$

where $u_i \overset{\text{i.i.d.}}{\sim} \gamma$ and $\epsilon_i \overset{\text{i.i.d.}}{\sim} \mu$

Goal

- Learn Ψ from samples $\mathcal{S}$

Method

- Deep neural networks $\Psi^n(u; \theta)$ as parametrization
- Supervised learning to find $\Psi^n(\cdot; \theta^*) \approx \Psi(\cdot)$

Literature

Existing theory

- A posteriori error analysis for DeepOnet[1]
- Non-DNN approach for linear operators[2]
- Approximation rates in applications[3]

Our goal

- A priori error analysis
- Nonlinear operators
- Discretization-invariant
- Curse of dimensionality in math applications

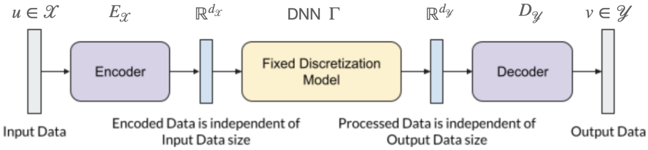
---

[1] S. Lanthaler, S. Mishra, and G. E. Karniadakis. Error estimates for deepOnets: A deep learning framework in infinite dimensions. arXiv:2102.09618, 2021.

[2] M. V. de Hoop, N. B. Kovachki, N. H. Nelsen, and A. M. Stuart. Convergence rates for learning linear operators from noisy data. arXiv:2108.12515, 2021.

[3] Deng et al, Approximation rates of DeepONets for learning operators arising from advection-diffusion equations, Neural Networks, Volume 153, September 2022, Pages 411-426

An Abstract Basic Framework



Encoder-decoder

- Most methods $\mathcal{X} \approx \mathbb{R}^{d_\mathcal{X}} \to \mathbb{R}^{d_\mathcal{Y}} \approx \mathcal{Y}$; finite basis expansion
- PCA-Net[4] $\mathcal{X} \approx \mathbb{R}^{d_\mathcal{X}} \to \mathbb{R}^{d_\mathcal{Y}} \approx \mathcal{Y}$; PCA
- DeepOnet $\mathcal{X} \approx \mathbb{R}^{d_\mathcal{X}} \to \mathcal{Y}$; $E_\mathcal{X}$: function sampling; $D_\mathcal{Y}$: DNN basis functions
- IAE-Net with one block, $\mathcal{X} \to \mathcal{Y}$; DNN kernels

[4] Bhattacharya, Hosseini, Kovachki, Stuart, 2019

Problem Statement

Learning $\Psi \approx D_{\mathcal{Y}} \circ \Gamma \circ E_{\mathcal{X}}$

- Target Lip. operator $\Psi : \mathcal{X} \to \mathcal{Y}$
- Samples $\mathcal{S} = \{u_i, v_i\}_{i=1}^{2n}$ with $v_i = \Psi(u_i) + \epsilon_i$, $u_i \overset{\text{i.i.d.}}{\sim} \gamma$, and $\epsilon_i \overset{\text{i.i.d.}}{\sim} \mu$
- Step 1: use $\{u_i, v_i\}_{i=1}^{n}$ to learn encoder-decoder s.t.

$$D_{\mathcal{X}} \circ E_{\mathcal{X}} \approx I \quad \text{and} \quad D_{\mathcal{Y}} \circ E_{\mathcal{Y}} \approx I$$

- Step 2: use $\{u_i, v_i\}_{i=n+1}^{2n}$ to learn DNN Γ_θ via empirical risk

$$\min_{\Gamma_\theta \in \mathcal{F}_{\mathrm{NN}}} R_S(\theta) := \min_{\Gamma_\theta \in \mathcal{F}_{\mathrm{NN}}} \frac{1}{n} \sum_{i=n+1}^{2n} \| D_{\mathcal{Y}} \circ \Gamma_\theta \circ E_{\mathcal{X}}(u_i) - v_i \|_{\mathcal{Y}}^2$$

- Population risk (accuracy) of $\Psi_\theta := D_{\mathcal{Y}} \circ \Gamma_\theta \circ E_{\mathcal{X}} \approx \Psi$

$$R_D(\theta) := \mathbb{E}_{\mathcal{S}} \mathbb{E}_{u \sim \gamma} \| \Psi_\theta(u) - \Psi(u) \|_{\mathcal{Y}}^2$$

Question: How good is the empirical solution Ψ_{θ^*} with $\theta^* \in \operatorname{argmin} R_S(\theta)$?

Problem Statement

The goal of error analysis

Quantify

$$R_D(\theta^*) := \mathbb{E}_{\mathcal{S}}\mathbb{E}_{u\sim\gamma}\|\Psi_{\theta^*}(u) - \Psi(u)\|_{\mathcal{Y}}^2$$

in terms of DNN width, depth, and $\#$samples

Key questions

- Practical guidance on the choice of DNNs and samples
- Curse of dimensionality (in $\#$parameters and $\#$samples)
- Zero/few-shot generalization for different data structures

Problem Statement

Error analysis of $R_D(\theta^*)$

- Error decomposition

$$
\begin{aligned}
R_D(\theta^*) &= \mathbb{E}_{\mathcal{S}}\mathbb{E}_{u\sim\gamma}\left[\|D_{\mathcal{Y}} \circ \Gamma_{\theta^*} \circ E_{\mathcal{X}}(u) - \Psi(u)\|_{\mathcal{Y}}^2\right] \\
&= T_1 + T_2
\end{aligned}
$$

- Bias (approximation)

$$
T_1 = 2\mathbb{E}_{\mathcal{S}}\left[\frac{1}{n}\sum_{i=n+1}^{2n}\|D_{\mathcal{Y}} \circ \Gamma_{\theta^*} \circ E_{\mathcal{X}}(u_i) - \Psi(u_i)\|_2^2\right]
$$

- Variance (generalization)

$$
\begin{aligned}
T_2 = &\ \mathbb{E}_{\mathcal{S}}\mathbb{E}_{u\sim\gamma}\left[\|D_{\mathcal{Y}} \circ \Gamma_{\theta^*} \circ E_{\mathcal{X}}(u) - \Psi(u)\|_{\mathcal{Y}}^2\right] \\
&- 2\mathbb{E}_{\mathcal{S}}\left[\frac{1}{n}\sum_{i=n+1}^{2n}\|D_{\mathcal{Y}} \circ \Gamma_{\theta^*} \circ E_{\mathcal{X}}(u_i) - \Psi(u_i)\|_2^2\right]
\end{aligned}
$$

First step: estimation of T_1 via DNN approximation

$$T_1 = 2\mathbb{E}_{\mathcal{S}}\left[\frac{1}{n}\sum_{i=n+1}^{2n}\|D_{\mathcal{Y}}\circ\Gamma_{\theta^*}\circ E_{\mathcal{X}}(u_i) - \Psi(u_i)\|_2^2\right]$$

Our goals in approximation

- Approximation error in terms of width and depth
- Does curse of dim (e.g., $\#$ parameters $(\frac{1}{\epsilon})^d$) exist?

Deep Network Approximation

Bias (approximation) with noise perturbation (np) and encoding error (ee)

$$
\begin{aligned}
T_1 &= 2\mathbb{E}_{\mathcal{S}}\left[\frac{1}{n}\sum_{i=n+1}^{2n}\|D_{\mathcal{Y}}\circ\Gamma_{\theta^*}\circ E_{\mathcal{X}}(u_i)-v_i\|_2^2\right]+\text{np}\\
&= 2\mathbb{E}_{\mathcal{S}}\left[\inf_{\Gamma\in\mathcal{F}_{\mathrm{NN}}}\frac{1}{n}\sum_{i=n+1}^{2n}\|D_{\mathcal{Y}}\circ\Gamma\circ E_{\mathcal{X}}(u_i)-v_i\|_2^2\right]+\text{np}\\
&\le 2\inf_{\Gamma\in\mathcal{F}_{\mathrm{NN}}}\mathbb{E}_{\mathcal{S}}\left[\frac{1}{n}\sum_{i=n+1}^{2n}\|D_{\mathcal{Y}}\circ\Gamma\circ E_{\mathcal{X}}(u_i)-v_i\|_2^2\right]+\text{np}\\
&= 2\inf_{\Gamma\in\mathcal{F}_{\mathrm{NN}}}\mathbb{E}_{\mu\sim\gamma}\left[\|D_{\mathcal{Y}}\circ\Gamma\circ E_{\mathcal{X}}(u)-\Psi(u)\|_2^2\right]+\text{np}\\
&= 2\inf_{\Gamma\in\mathcal{F}_{\mathrm{NN}}}\mathbb{E}_{\mu\sim\gamma}\left[\|D_{\mathcal{Y}}\circ\Gamma\circ E_{\mathcal{X}}(u)-D_{\mathcal{Y}}\circ E_{\mathcal{Y}}\circ\Psi(u)\|_2^2\right]+\text{np}+\text{ee}\\
&\le 2L_{\mathcal{Y}}^2\inf_{\Gamma\in\mathcal{F}_{\mathrm{NN}}}\mathbb{E}_{\mu\sim\gamma}\left[\|\Gamma\circ E_{\mathcal{X}}(u)-E_{\mathcal{Y}}\circ\Psi(u)\|_2^2\right]+\text{np}+\text{ee},
\end{aligned}
$$

which is a DNN regression problem from $E_{\mathcal{X}}(u)\in\mathbb{R}^{d_{\mathcal{X}}}$ to $E_{\mathcal{Y}}\circ\Psi(u)\in\mathbb{R}^{d_{\mathcal{Y}}}$

Deep Network Approximation

Bias (approximation) with noise perturbation (np) and encoding error (ee)

$$T_1 \le 2 \inf_{\Gamma \in \mathcal{F}_{\mathrm{NN}}} \mathbb{E}_{\mu\sim\gamma}\left[\| D_{\mathcal{Y}} \circ \Gamma \circ E_{\mathcal{X}}(u) - D_{\mathcal{Y}} \circ E_{\mathcal{Y}} \circ \Psi \circ D_{\mathcal{X}} \circ E_{\mathcal{X}}(u) \|_2^2\right] + \mathrm{np} + \mathrm{ee}$$

$$\le 2L_{\mathcal{Y}}^2 \inf_{\Gamma \in \mathcal{F}_{\mathrm{NN}}} \mathbb{E}_{\mu\sim\gamma}\left[\| \Gamma\left(E_{\mathcal{X}}(u)\right) - E_{\mathcal{Y}} \circ \Psi \circ D_{\mathcal{X}}\left(E_{\mathcal{X}}(u)\right) \|_2^2\right] + \mathrm{np} + \mathrm{ee},$$

- DNN approximation to $E_{\mathcal{Y}} \circ \Psi \circ D_{\mathcal{X}}$ from $\mathbb{R}^{d_{\mathcal{X}}}$ to $\mathbb{R}^{d_{\mathcal{Y}}}$

Our goals in approximation

- Approximation error in terms of width and depth
- Does curse of dim (e.g., $\#$ parameters $(\frac{1}{\epsilon})^d$) exist?

Literature Review

Active research directions

Cybenko, 1989; Hornik et al., 1989; Barron, 1993; Montufar, Ay, 2011; Liang and Srikant, 2016; Yarotsky, 2017; Poggio et al., 2017; Schmidt-Hieber, 2017; E and Wang, 2018; Petersen and Voigtlaender, 2018; Chui et al., 2018; Yarotsky, 2018; Nakada and Imaizumi, 2019; Gribonval et al., 2019; Gühring et al., 2019; Chen et al., 2019; Li et al., 2019; Suzuki, 2019; Bao et al., 2019; E et al., 2019; Opschoor et al., 2019; Merkh, Montufar, 2019; Yarotsky and Zhevnerchuk, 2019; Bölcskei et al., 2019; Montanelli and Du, 2019; Chen and Wu, 2019; Zhou, 2020; Montanelli et al., 2020, etc.

ReLU DNNs, continuous functions $C([0,1]^d)$

ReLU; Fixed network width $O(N)$ and depth $O(L)$

- Nearly tight error rate $5\omega_f(8\sqrt{d}N^{-2/d}L^{-2/d})$ simultaneously in N and L with L^∞-norm. Shen, Y., and Zhang (CiCP, 2020)
- ω_f is the modulas of continuity
- Improved to a tight rate $O\left(\sqrt{d}\,\omega_f\left(\left(N^2L^2\log_3(N+2)\right)^{-1/d}\right)\right)$. Shen, Y., and Zhang (J Math Pures Appl, 2021)

Remark

- Curse of dim exists
- Smoothness cannot help (Lu, Shen, Y., Zhang, SIMA, 2021)
- Need special function structures or activation functions to lessen the curse

DNNs with advanced activation function

Elementary universal activation function (EUAF)

$$\sigma_1(x) = \left|x - 2\lfloor \tfrac{x+1}{2} \rfloor\right|,$$

$$\sigma_2(x) := \frac{x}{|x|+1},$$

$$\sigma(x) := \begin{cases} \sigma_1(x) & \text{for } x \in [0, \infty), \\ \sigma_2(x) & \text{for } x \in (-\infty, 0). \end{cases}$$

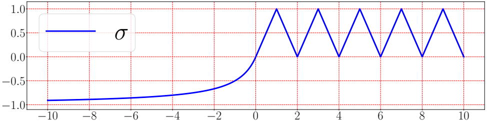


Figure: An illustration of σ on $[-10, 10]$.

DNNs with advanced activation function

Theorem (EUAF approximation in *d*-dimensions)

Arbitrarily small error with a fixed number of neurons for $C([0,1]^d)$.

- *For any* $\epsilon > 0$, *there exists* ϕ *of width* $36d(2d+1)$ *and depth* 11 *s.t.*

$$\|f(x) - \phi(x)\|_{L^\infty([0,1]^d)} \leq \epsilon$$

- *Shen, Y., and Zhang (arXiv:2107.02397)*

Remark

- No curse of dim in approximation
- But the pseudo-dimension of these DNNs is infinite

Second step: estimation of T_2 via DNN generalization

$$
\begin{aligned}
T_2 = {} & \mathbb{E}_{\mathcal{S}} \mathbb{E}_{u\sim\gamma} \left[\| D_{\mathcal{Y}} \circ \Gamma_{\theta^*} \circ E_{\mathcal{X}}(u) - \Psi(u) \|_{\mathcal{Y}}^2 \right] \\
& - 2\mathbb{E}_{\mathcal{S}} \left[\frac{1}{n} \sum_{i=n+1}^{2n} \| D_{\mathcal{Y}} \circ \Gamma_{\theta^*} \circ E_{\mathcal{X}}(u_i) - \Psi(u_i) \|_2^2 \right]
\end{aligned}
$$

Deep Network Generalization

Active research directions

Hamers and Kohler 2006; Jacot, Gabriel, and Hongler 2018; Bauer and Kohler 2019; Cao and Gu 2019; Chen et al. 2019; Schmidt-Hieber 2020; Kohler, Krzyzak, and Langer 2020; Nakada and Imaizumi 2020; Farrell, Liang, and Misra 2021; Jiao, Shen, Lin, and Huang 2021, etc

Remark

Very limited for operator learning

Deep Network Generalization

Road map (Liu, Y.*, Chen, Zhao, Liao*, arXiv:2201.00217, 2022)

- Variance $T_2 \to$ covering number of $\mathcal{F}_{\mathrm{NN}}$
- Covering number of $\mathcal{F}_{\mathrm{NN}} \to$ pseudo-dimension of $\mathcal{F}_{\mathrm{NN}}$
- Pseudo-dimension of $\mathcal{F}_{\mathrm{NN}} \to$ NN width and depth

Deep Network Generalization

Road map

Variance T_2 $\to$ covering number of $\mathcal{F}_{\rm NN}$ $\to$ pseudo-dimension of $\mathcal{F}_{\rm NN}$

Variance bound (Liu, Y.*, Chen, Zhao, Liao*, arXiv:2201.00217, 2022)

$$
\begin{aligned}
T_2 = & \mathbb{E}_{\mathcal{S}}\mathbb{E}_{u\sim\gamma}\left[\|D_{\mathcal{Y}}\circ\Gamma_{\theta^*}\circ E_{\mathcal{X}}(u)-\Psi(u)\|_{\mathcal{Y}}^2\right] \\
& -2\mathbb{E}_{\mathcal{S}}\left[\frac{1}{n}\sum_{i=n+1}^{2n}\|D_{\mathcal{Y}}\circ\Gamma_{\theta^*}\circ E_{\mathcal{X}}(u_i)-\Psi(u_i)\|_2^2\right] \\
\leq & O\left(\frac{1}{n}\log\mathcal{N}\left(O(\delta),\mathcal{F}_{\rm NN},2n\right)\right)+O(\delta),
\end{aligned}
$$

where $\mathcal{N}\left(O(\delta),\mathcal{F}_{\rm NN},2n\right)$ is the δ-covering number of $\mathcal{F}_{\rm NN}$ at $2n$ samples

Deep Network Generalization

Road map

Variance $T_2 \to$ covering number of $\mathcal{F}_{\rm NN} \to$ pseudo-dimension of $\mathcal{F}_{\rm NN}$

$$
\begin{aligned}
T_2 &\leq O\left(\frac{1}{n}\log\mathcal{N}\left(O(\delta),\mathcal{F}_{\rm NN},2n\right)\right)+O(\delta)\\
&\leq O\left(\frac{1}{n}\mathrm{Pdim}(\mathcal{F}_{\rm NN})\log\left(O\left(\frac{n}{\delta\mathrm{Pdim}(\mathcal{F}_{\rm NN})}\right)\right)\right)+O(\delta)
\end{aligned}
$$

Theorem (Anthony and Bartlett, 1999)

Let F be a class of functions from some domain Ω to $[-M, M]$. Denote the pseudo-dimension of F by $\mathrm{Pdim}(F)$. For any $\delta > 0$, we have

$$
\mathcal{N}(\delta, F, 2n) \leq \left(\frac{4eMn}{\delta\mathrm{Pdim}(F)}\right)^{\mathrm{Pdim}(F)}
$$

for $2n > \mathrm{Pdim}(F)$.

Deep Network Generalization

Road map

Variance $T_2 \to$ covering number of $\mathcal{F}_{\mathrm{NN}} \to$ pseudo-dimension of $\mathcal{F}_{\mathrm{NN}}$

$$
\begin{aligned}
T_2 &\leq O\left(\frac{1}{n}\mathrm{Pdim}(\mathcal{F}_{\mathrm{NN}}) \log\left(O\left(\frac{n}{\delta \mathrm{Pdim}(\mathcal{F}_{\mathrm{NN}})}\right)\right)\right) + O(\delta) \\
&\leq O\left(\frac{1}{n} LU \log(U) \log\left(O\left(\frac{n}{\delta LU \log(U)}\right)\right)\right) + O(\delta),
\end{aligned}
$$

where $U = N^2 L$ for DNN of width N and depth L

Theorem (Bartlett, Harvey, Liaw, and Mehrabian, 2019)

For ReLU network architecture $\mathcal{F}_{\mathrm{NN}}$ with L layers and U parameters, there exists a universal constant C such that

$$\mathrm{Pdim}(\mathcal{F}_{\mathrm{NN}}) \leq CLU \log(U).$$

Full Error Analysis

Theorem ((Liu, Y.*, Chen, Zhao, Liao*, arXiv:2201.00217))

Under certain assumptions. Let Γ_{θ^*} be the minimizer of the empirical loss with depth $L = O(\widetilde{L}\log\widetilde{L})$, width $N = O\left(\widetilde{p}\log\widetilde{p}\right)$, magnitude bound $M = O(\sqrt{d_{\mathcal{Y}}})$, where $\widetilde{L}, \widetilde{p}$ are positive integers satisfying

$$\widetilde{L}\widetilde{p} = \left\lceil d_{\mathcal{Y}}^{-\frac{d_{\mathcal{X}}}{4+2d_{\mathcal{X}}}} n^{\frac{d_{\mathcal{X}}}{4+2d_{\mathcal{X}}}} \right\rceil .$$

Then we have

$$\begin{aligned}
&\mathbb{E}_{\mathcal{S}}\mathbb{E}_{u\sim\gamma}\|D_{\mathcal{Y}}\circ\Gamma_{\theta^*}\circ E_{\mathcal{X}}(u)-\Psi(u)\|_{\mathcal{Y}}^2\\
\leq\, & O\left((\sigma^2+1)d_{\mathcal{Y}}^{\frac{4+d_{\mathcal{X}}}{2+d_{\mathcal{X}}}} n^{-\frac{2}{2+d_{\mathcal{X}}}}\log^6 n\right)\\
&\quad + O\left(\mathbb{E}_{\mathcal{S}}\mathbb{E}_{u\sim\gamma}\|D_{\mathcal{X}}\circ E_{\mathcal{X}}(u)-u\|_{\mathcal{X}}^2+\mathbb{E}_{\mathcal{S}}\mathbb{E}_{w\sim\Psi_{\#}\gamma}\|D_{\mathcal{Y}}\circ E_{\mathcal{Y}}(w)-w\|_{\mathcal{Y}}^2\right)
\end{aligned}$$

Interpretation

- Curse of dim exists
- Require accurate encoding for zero/few-shot generalization

Additional Low-Dimensional Structures

Assumption (low-dimensional manifold)

$\{E_{\mathcal{X}}(u) : u \sim \gamma\}$ lie on a d_0-dimensional manifold with $d_0 \ll d_{\mathcal{X}}$

Theorem ((Liu, Y.*, Chen, Zhao, Liao*, arXiv:2201.00217))

In addition to the above assumption, we have

$$\begin{aligned}
&\mathbb{E}_{\mathcal{S}}\mathbb{E}_{u\sim\gamma}\|D_{\mathcal{Y}} \circ \Gamma_{\theta^*} \circ E_{\mathcal{X}}(u) - \Psi(u)\|_{\mathcal{Y}}^2 \\
&\leq O\left((\sigma^2+1) d_{\mathcal{Y}}^{\frac{4+d_0}{2+d_0}} n^{-\frac{2}{2+d_0}} \log^6 n\right) \\
&\quad + O\left(\mathbb{E}_{\mathcal{S}}\mathbb{E}_{u\sim\gamma}\|D_{\mathcal{X}} \circ E_{\mathcal{X}}(u) - u\|_{\mathcal{X}}^2 + \mathbb{E}_{\mathcal{S}}\mathbb{E}_{w\sim\Psi_{\#}\gamma}\|D_{\mathcal{Y}} \circ E_{\mathcal{Y}}(w) - w\|_{\mathcal{Y}}^2\right)
\end{aligned}$$

Additional Low-Dimensional Structures

Assumption (low complexity)

$$D_{\mathcal{Y}} \circ E_{\mathcal{Y}} \circ \Psi(u) = D_{\mathcal{Y}} \circ \mathbf{g} \circ E_{\mathcal{X}}(u)$$

with $\mathbf{g} : \mathbb{R}^{d_{\mathcal{X}}} \to \mathbb{R}^{d_{\mathcal{X}}}$ in the form:

$$\mathbf{g}(\mathbf{a}) = \begin{bmatrix} g_1(V_1^\top \mathbf{a}) & \cdots & g_{d_{\mathcal{Y}}}(V_{d_{\mathcal{Y}}}^\top \mathbf{a}) \end{bmatrix}^\top,$$

for $V_k \in \mathbb{R}^{d_{\mathcal{X}} \times d_0}$, and $g_k : \mathbb{R}^{d_0} \to \mathbb{R}$ (multi-index models).

Theorem ((Liu, Y.*, Chen, Zhao, Liao*, arXiv:2201.00217))

In addition to the above assumption, we have

$$\begin{aligned}
&\mathbb{E}_{\mathcal{S}}\mathbb{E}_{u\sim\gamma}\|D_{\mathcal{Y}} \circ \Gamma_{\theta^*} \circ E_{\mathcal{X}}(u) - \Psi(u)\|_{\mathcal{Y}}^2 \\
&\le O\left((\sigma^2+1) d_{\mathcal{Y}}^{\frac{4+d_0}{2+d_0}} \max\left\{n^{-\frac{2}{2+d_0}}, d_{\mathcal{X}} n^{-\frac{4+d_0}{4+2d_0}}\right\} \log^6 n\right) \\
&\quad + O\left(\mathbb{E}_{\mathcal{S}}\mathbb{E}_{u\sim\gamma}\|D_{\mathcal{X}} \circ E_{\mathcal{X}}(u) - u\|_{\mathcal{X}}^2 + \mathbb{E}_{\mathcal{S}}\mathbb{E}_{w\sim\Psi_{\#}\gamma}\|D_{\mathcal{Y}} \circ E_{\mathcal{Y}}(w) - w\|_{\mathcal{Y}}^2\right)
\end{aligned}$$

Our goal

- A priori error analysis
- Nonlinear operators
- Discretization-invariant
- Curse of dimensionality in math applications

No Curse Conditions

Learning error:

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{S}}\mathbb{E}_{u\sim\gamma}\|D_{\mathcal{Y}}\circ\Gamma_{\theta^*}\circ E_{\mathcal{X}}(u)-\Psi(u)\|_{\mathcal{Y}}^2\\
\leq\ & O\left((\sigma^2+1)d_{\mathcal{Y}}^{\frac{4+d_{\mathcal{X}}}{2+d_{\mathcal{X}}}}n^{-\frac{2}{2+d_{\mathcal{X}}}}\log^6 n\right)\\
&\quad+O\left(\mathbb{E}_{\mathcal{S}}\mathbb{E}_{u\sim\gamma}\|D_{\mathcal{X}}\circ E_{\mathcal{X}}(u)-u\|_{\mathcal{X}}^2+\mathbb{E}_{\mathcal{S}}\mathbb{E}_{w\sim\Psi_{\#}\gamma}\|D_{\mathcal{Y}}\circ E_{\mathcal{Y}}(w)-w\|_{\mathcal{Y}}^2\right)
\end{aligned}
$$

- Operator has low-dimensional or low-complexity structures
- No curse in the encoders for $\mathcal{X}$ and $\mathcal{Y}$

Question: What math applications satisfy these conditions?

Example 1: Burgers equation

$$\partial_t u(x,t) + \partial_x(u^2(x,t)/2) = \nu \partial_{xx} u(x,t), \quad x \in (0,1), t \in (0,1]$$
$$u(x,0) = u_0(x)$$

- Periodic boundary conditions
- $\nu = 0.1$: a given viscosity coefficient
- Applications in fluid mechanics, nonlinear acoustics, gas dynamics, and traffic flow
- Goal: learn the mapping from $u_0(x)$ to $u(x, T)$.

Example: Burgers Equation

- $\mathcal{X}$ and $\mathcal{Y}$ are spaces of 1D periodic function
- For Lip. continuous functions, K-dimensional vector space with $O(1/K)$ encoder errors

Example: Burgers Equation

The solution map can be explicitly written using the Cole-Hopf transformation $u = \frac{-2\nu v_x}{v}$ where

$$\begin{cases} v_t = \nu v_{xx} \\ v(x,0) = v_0(x) = \exp\left(-\frac{1}{2\nu}\int_{-\pi}^{x} u_0(s)ds\right) \end{cases} \tag{1}$$

and the solution is given by

$$u(x,T) = -2\nu \frac{\int_{\mathbb{R}} \partial_x \mathcal{K}(x,y,T) v_0(y) dy}{\int_{\mathbb{R}} \mathcal{K}(x,y,T) v_0(y) dy}$$

where the integration kernel $\mathcal{K}$ is defined as $\mathcal{K}(x,y,t) = \frac{1}{\sqrt{4\pi\nu t}} \exp\left(\frac{-(x-y)^2}{4\pi t}\right)$. Therefore, the solution map has low-complexity.