NEURAL INFORMATION PROCESSING SYSTEMS

Understanding Generalization in Physics Informed Models through Affine Variety Dimensions

Takeshi Koshizuka (The University of Tokyo)

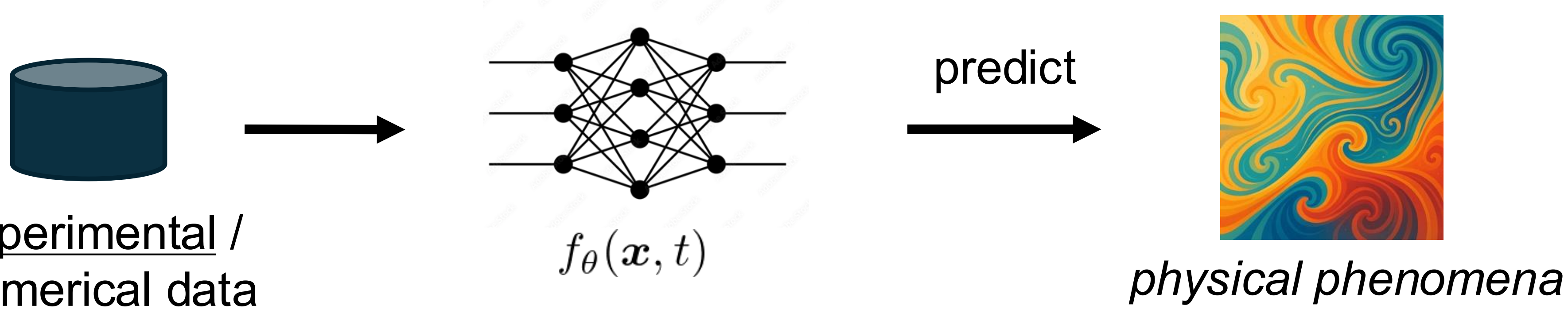
Issei Sato (The University of Tokyo)

UTokyo

Background: Physics-informed Machine Learning

What is PIML?

- Combine experimental/numerical data and physical constraints
- Incorporate physics laws into model architecture or loss function, etc.



Examples of Physics laws

Differential Equations	Symmetries	Stability
$\partial_t f_\theta = \mathcal{A} f_\theta$	$f_\theta(g \cdot x) = g \cdot f_\theta(x)$	$\lim_{t\to\infty} f_\theta = f^*$

- Empirically improves generalization & OOD robustness
- But theoretical understanding is limited

The goal is to provide a unified theoretical framework explaining how physics constraints improve generalization.

Formulation: Complexity Error Analysis on PILR

Data

- **Target function:** f^* s.t. $\mathscr{D}[f^*] \simeq 0$
- **Training samples:** $\{(x_i, y_i)\}_{i=1}^{n}$
- **Observation model:**

$$y_i = f^*(x_i) + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Physics-informed Linear Regression (PILR)

$$\hat{\boldsymbol{w}} = \underset{\boldsymbol{w} \in \mathcal{V}(\mathscr{D}, \mathcal{B}, \mathcal{T}) \cap \mathbb{B}_2(R)}{\arg\min} \frac{1}{n} \| \boldsymbol{y} - \boldsymbol{\Phi} \boldsymbol{w} \|_2^2$$

Feasible Set Definition

$$\mathcal{V}(\mathscr{D}, \mathcal{B}, \mathcal{T}) := \left\{ \boldsymbol{w} \in \mathbb{R}^d : \langle \mathscr{D}[\boldsymbol{w}^\top \phi], \psi_k \rangle_{\mu_k} = 0, \forall (\psi_k, \mu_k) \in \mathcal{T}, \phi_j \in \mathcal{B} \right\}$$

	Collocation	**Variational**
Measure μ_k	Dirac δ_{x_k}	Lebesgue/Borel on Ω
Trial ψ_k	constant or probes ($\psi_k \equiv 1$)	FE basis/test space V_h
Enforcement	pointwise residual $\mathscr{D}[f](x_k) = 0$	weak residual $\int \mathscr{D}[f] \psi_k \, dx = 0$
Pros	simple, mesh-free	stability, boundary handling, FEM links

Min-Max Complexity Error Analysis

$$\mathcal{E}_{\mathrm{PILR}} = \min_{\widehat{\boldsymbol{w}}} \max_{\boldsymbol{w}^* \in \mathcal{V} \cap \mathbb{B}_2(R)} \| \widehat{\boldsymbol{w}} - \boldsymbol{w}^* \|_2^2$$

Represent physics constraints via a *unified residual form*, integrating **collocation (PINNs)** and **variational (FEM)** approaches.

Main Results

Regularity Assumptions

1. **Bounded basis:** Each basis function $\phi_j \in \mathcal{B}$ is bounded, i.e., $\sup_{x\in\Omega} \|\phi(x)\|_2 \leq M_\phi$.
2. **Well-conditioned design:** For some $\eta > 0$, the design matrix $\boldsymbol{\Phi}$ satisfies

$$\frac{1}{\sqrt{n}}\|\boldsymbol{\Phi w}\|_2 \geq \frac{1}{\sqrt{\eta}}\|\boldsymbol{w}\|_2, \quad \forall\, \boldsymbol{w} \in \mathbb{B}_2(2R),$$

3. **Estimator stability:** The estimator is Lipschitz-continuous w.r.t. the optimal weights:

$$\|\hat{\boldsymbol{w}}_1 - \hat{\boldsymbol{w}}_2\|_2 \leq (\Gamma - 1)\|\boldsymbol{w}_1^* - \boldsymbol{w}_2^*\|_2, \quad \Gamma > 1.$$

Theorem (informal)

Let $\mathcal{V}(\mathscr{D}, \mathcal{B}, \mathcal{T})$ be the $(\beta, d_{\mathcal{V}})$-regular affine variety. Suppose that

- the basis functions are bounded by a constant,
- the minimum eigenvalue of the design matrix is restricted, and
- the stability condition for the estimator holds.

Then, for $\delta \in (0, 1)$, with probability at least $1 - \delta$, the minimax risk for PILR satisfies

$$\mathcal{O}\left(\sqrt{\frac{d_{\mathcal{V}} \log(d_{\mathcal{V}} d)}{n}} + \sqrt{\frac{\log 2\beta}{n}} + 2\sqrt{\frac{\log(2/\delta)}{n}}\right).$$

Main Insight

- The minimax risk depends on the **intrinsic dimension** $d_{\mathcal{V}}$, not the ambient dimension d.
- When the topological complexity β is small:

$$\mathcal{E}_{\text{PILR}} = \mathcal{O}\left(\sqrt{\frac{d_{\mathcal{V}} \log d_{\mathcal{V}} d}{n}}\right) \quad \text{vs.} \quad \mathcal{E}_{\text{RR}} = \mathcal{O}\left(\sqrt{\frac{d}{n}}\right).$$

- Embedding physical constraints $\Rightarrow$ reduced effective hypothesis complexity.

Role of β

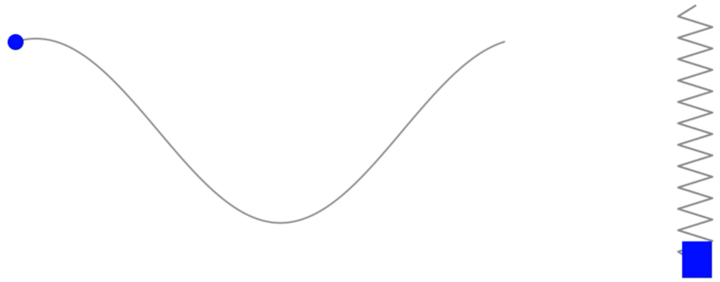
- β measures the **topological complexity** of the affine variety (sum of Betti numbers). If the operator $\mathscr{D}$ is linear, $\beta = 1$.
- Upper bound via **Petrovskii–Oleinik–Milnor inequality:**

$$\beta \leq \rho(2\rho - 1)^{d+1},$$

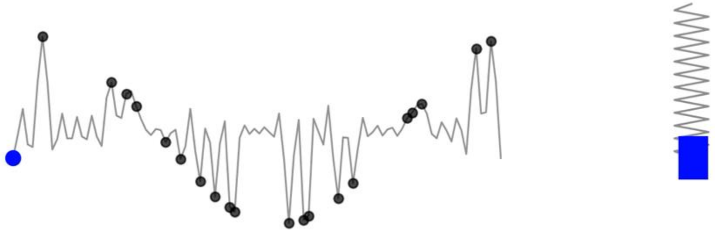
where ρ is the maximal polynomial degree.
- Larger $\rho \Rightarrow$ more holes and disconnected components.

Example

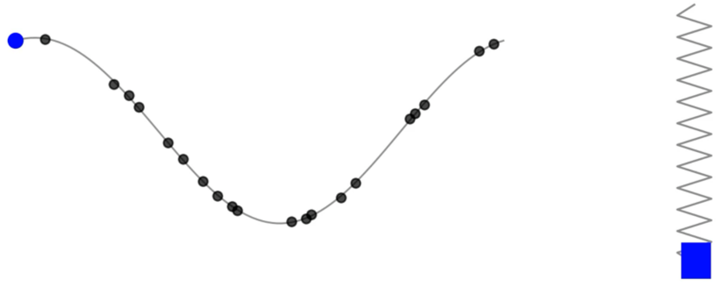
Problem: Harmonic Oscillator



Linear Regression w/ Fourier Basis



Physics-informed Linear Regression



Equation

$$\mathscr{D}[y] = \frac{\mathrm{d}^2}{\mathrm{d}t^2} y + \frac{k_s}{m_s} y,$$

Analytical solution

$$y(t) = y_0 \cos(\omega t) + \frac{v_0}{\omega} \sin(\omega t), \omega = \sqrt{k_s/m_s}.$$

Basis/Trial Functions

$$\{\phi_j(x)\} = \{1, \cos(\omega_j x), \sin(\omega_j x)\},$$

$$\psi_k(x) = 1,\ \ \mu_k = \delta_{x_k}$$

(Dirac measure)

Complexity error bound

$$\mathcal{O}\left(\sqrt{\frac{d}{n}}\right)$$

When d (number of basis) $\gg$ n (sample size), performance degrades due to overfitting.

vs

$$\mathcal{O}\left(\sqrt{\frac{d_{\mathcal{V}} \log d_{\mathcal{V}} d}{n}}\right)$$

When $\mathrm{d} \gg \mathrm{d}_{\mathcal{V}}$ (the dimension of the affine variety), performance improves.

$$d_{\mathcal{V}} = \dim\left(\{\mathbf{w} \in \mathbb{R}^d : \mathscr{D}[\mathbf{w}^\top \phi](x_k) = 0,\ \forall k\}\right) = 2$$

= number of basis : $\cos(\omega t)$ and $\sin(\omega t)$