Physics-informed machine learning as a kernel method

$\mathbb{R}^d$ out OK but tedious
trivial extension when the PDE does not involve cross terms in the coordinates
↳ separable PDE

- Hybrid modeling

<u>Model</u>; $Y = f^*(X) + \varepsilon$ $\quad f^*: \mathbb{R}^d \to \mathbb{R}$

$X \in \Omega \subseteq [-L, L]^d$

input domain as a bounded Lipschitz domain

includes $\mathcal{C}^1$-manifolds (e.g. the Euclidean ball $\{x \in \mathbb{R}^d \; \|x\| \leq 1\}$

domains with non-differentiable boundaries (e.g. $[-L, L]^d$).

<u>Hyp</u>: $f^* \in H^s(\Omega)$ for some $s > d/2$.

$\mathscr{D}(f^*) \simeq 0$ for some <u>known</u> differential operator.

<u>Empirical risk function</u>:

hyperparameters

$$R_n(f) = \frac{1}{n} \sum_{i=1}^{n} |f(X_i) - Y_i|^2 + \lambda_{(sob)} \|f\|^2_{H^s_{per}([-2L; 2L]^d)} + \lambda_{(pde)} \|\mathscr{D}(f)\|^2_{L^2(\Omega)}$$

Minimization over the class $H^s_{per}([-2L; 2L]^d)$ = subspace of $H^s([-2L; 2L]^d)$ consisting of functions whose $4L$-periodic extension is still s-times weakly differentiable.

<u>Any</u> function $f \in H^s(\Omega)$ can be extended to a function in $H^s_{per}([-2L, 2L]^d)$.

$$H^s(\Omega) \hookrightarrow H^s_{per}([-2L, 2L]^d)$$

Study of the estimator

$$\hat{f}_n = \underset{f \in H^s_{per}([-2L,2L]^d)}{\operatorname{argmin}} R_n(f) \ \circledast$$

$\circledast$ All derived results remain applicable to the standard Sobolev space $H^s(\Omega)$. But considering $H^s_{per}([-2L,2L]^d)$ eases technicalities in the proofs.

Goal:

Framing PIML as a kernel method

$$\hat{f}_n = \underset{f \in H^s_{\mathrm{per}}([-2L,2L]^d)}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n} |f(X_i) - Y_i|^2 + \|f\|^2_{\mathrm{RKHS}},$$

with $\|f\|^2_{\mathrm{RKHS}} = \lambda_{(\mathrm{sob})} \|f\|^2_{H^s_{\mathrm{per}}([-2L,2L]^d)} + \lambda_{(\mathrm{pde})} \|\mathscr{D}(f)\|^2_{L^2(\Omega)}$

- How does the PDE penalty impact learning?
- How to leverage the kernel toolbox?
- How to define a tractable estimator?

Hyp: The operator $\mathscr{D} : H^s(\Omega) \to L^2(\Omega)$ is assumed to be a linear differential operator if for all $f \in H^s(\Omega)$

$$\mathscr{D}(f) = \sum_{|\alpha| \leq s} p_\alpha \, \partial^\alpha f$$

$p_\alpha : \Omega \to \mathbb{R}$ functions

s.t. $\max_\alpha \|p_\alpha\|_\infty < +\infty$

- <u>First step</u>: Any function $f \in H^s_{per}([-2L,2L]^d)$ can be linearly mapped in $L^2([-2L,2L]^d)$ in such a way that the norm $\|.\|_{L^2([-2L,2L]^d)}$ of the embedding is equal to $\lambda_{(sob)}\|f\|^2_{H^s_{per}([-2L,2L]^d)} + \lambda_{(pde)}\|\mathscr{D}(f)\|^2_{L^2(\Omega)}$

Lemma

There exists a positive operator $\mathscr{O}_n$ on $L^2([-2L,2L]^d)$ such that, for any $f \in H^s_{\text{per}}([-2L,2L]^d)$,

$$\|\mathscr{O}_n^{-1/2}(f)\|^2_{L^2([-2L,2L]^d)} = \lambda_{(\text{sob})}\|f\|^2_{H^s_{\text{per}}([-2L,2L]^d)} + \lambda_{(\text{pde})}\|\mathscr{D}(f)\|^2_{L^2(\Omega)}.$$

Remark: $\mathscr{O}_n^{-1/2} : H^s_{per}([-2L,2L]^d) \to L^2([-2L,2L]^d)$

<u>Ideas of proof</u>:

$\exists$ injective operator $\mathscr{O}_n : L^2([-2L,2L]^d) \to H^s_{per}([-2L,2L]^d)$

s.t $\forall f \in L^2([-2L,2L]^d)$, $\mathscr{O}_n(f)$ is the unique elt of $H^s_{per}([-2L,2L]^d)$ solution of the weak formulation:

$$\forall \phi \in H^s_{per}([-2L,2L]^d),\ \underbrace{\lambda_{(sob)} \sum_{|\alpha|\leq s} \int_{[-2L,2L]^d} \partial^\alpha \phi\, \partial^\alpha \mathscr{O}_n(f) + \lambda_{(pde)} \int_\Omega \mathscr{D}\phi \mathscr{D}\mathscr{O}_n(f)}_{=: B(\phi,\, \mathscr{O}_n(f))} = \int_{[-2L,2L]^d} \phi f$$

$B(u,u) \geq \lambda_{(sob)} \|u\|^2_{H^s_{per}}$

$\hookrightarrow$ $B(u,v)$ is a bilinear form on $H^s_{per}([-2L,2L]^d)$ + coercive

continuous $|B(u,v)| \leq (\lambda_{(sob)} + \ldots)\|u\|_{H^s_{per}} \|v\|_{H^s_{per}}$

$\phi \mapsto \int \phi f$ is a bounded linear form on H^s_{per}.

$\hookrightarrow$ Lax-Milgram thm: $\forall f \in L^2([-2L,2L]^d)$, $\exists! w \in H^s_{per}$ s.t $\forall \phi \in H^s_{per}$ $B(\phi, w) = \int \phi f$.

$\hookrightarrow$ Call $\mathscr{O}_n : f \mapsto \mathscr{O}_n(f) = w$. (injective thanks to <u>uniqueness</u>)

↳ $\|G_n(f)\|_{H^s_{per}} \leqslant \lambda^{-1}_{(sob)} \|f\|_{L^2([-2L,2L]^d)}$.

Indeed,

$$\|G_n f\|^2_{H^s_{per}} \leqslant \lambda^{-1}_{(sob)} B[G_n(f), G_n(f)] = \lambda^{-1}_{(sob)} \langle G_n(f), f \rangle_{L^2([-2L,2L]^d)}$$
$$\leqslant \lambda^{-1}_{(sob)} \|f\|_{L^2} \|G_n(f)\|_{L^2}$$
$$\leqslant \lambda^{-1}_{(sob)} \|f\|_{L^2} \|G_n(f)\|_{H^s_{per}}$$

orthonormal basis of $L^2([-2L,2L]^d)$.

↳ $G_n = \sum_{m \in \mathbb{N}} a_m \langle v_m, \cdot \rangle_{L^2([-2L,2L]^d)} v_m$

G_n can be diagonalized thanks to the spectral thm, G_n being a compact operator (Rellich – Kondrakov) and strictly positive ($\langle f, G_n(f) \rangle_{L^2} = B[G_n(f), G_n(f)] \geqslant \lambda_{(sob)} \|f\|_{H^s_{per}} > 0$).

↳ $(v_m)_m$ is also an orthonormal basis of H^s_{per}.

↳ $G_n^{1/2} : L^2([-2L,2L]^d) \longrightarrow H^s_{per}([-2L,2L]^d)$ is well-defined.

$G_n^{-1/2} : H^s_{per}([-2L,2L]^d) \longrightarrow L^2([-2L,2L]^d)$ is well defined.

$$G_n^{-1/2} = \sum_{m \in \mathbb{N}} a_m^{-1/2} \langle v_m, \cdot \rangle_{L^2([-2L,2L]^d)} v_m .$$

↳ $G_n^{1/2}(\delta_x) \in L^2([-2L,2L]^d)$

This suggests the inner product "$\langle f,g\rangle$" $= \langle \mathcal{O}_n^{-1/2}(f), \mathcal{O}_n^{-1/2}(g)\rangle_{L^2([-2L,2L]^d)}$

More intuition

Morally,

$$f(x) \text{ `` = '' } \langle f, \delta_x\rangle = \langle \mathscr{O}_n^{-1/2}(f), \mathscr{O}_n^{1/2}(\delta_x)\rangle_{L^2([-2L,2L]^d)}$$

One can recognize here a reproducing property

$$f(x) = \langle f, \mathscr{O}_n(\delta_x)\rangle,$$

where the inner product would be given by

$$\langle g, h\rangle = \langle \mathscr{O}_n^{-1/2}(g), \mathscr{O}_n^{-1/2}(h)\rangle_{L^2([-2L,2L]^d)}.$$

- RKHS structure :

- For any $f \in L^2([-2L,2L]^d)$ and $x \in [-2L,2L]^d$,

$$\mathscr{O}_n(f)(x) = \sum_{m\in\mathbb{N}} a_m \langle f, v_m\rangle_{L^2([-2L,2L]^d)} v_m(x)$$

- Orthonormal basis of eigenfunctions $v_m \in H^s_{\mathrm{per}}([-2L,2L]^d)$
- Eigenvalues $a_m > 0$

Theorem

The space $H^s_{\mathrm{per}}([-2L,2L]^d)$, equipped with the inner product

$$\langle f, g\rangle_{\mathrm{RKHS}} = \langle \mathscr{O}_n^{-1/2} f, \mathscr{O}_n^{-1/2} g\rangle_{L^2([-2L,2L]^d)},$$

is a reproducing kernel Hilbert space. For $f \in H^s_{\mathrm{per}}([-2L,2L]^d)$,

$$\|f\|^2_{\mathrm{RKHS}} = \lambda_{(\mathrm{sob})}\|f\|^2_{H^s_{\mathrm{per}}([-2L,2L]^d)} + \lambda_{(\mathrm{pde})}\|\mathscr{D}(f)\|^2_{L^2(\Omega)},$$

and the associated kernel is $K(x,y) = \sum_{m\in\mathbb{N}} a_m v_m(x) v_m(y)$.

Proof (sketch)

- take $\Omega = [-\pi,\pi]^d = [-2L,2L]^d$ with $L = \pi/2$ (functions with periodic derivatives on Ω)
- Assume that $\mathscr{D}$ has constant coefficients (penalized by the PDE on the whole domain).

- In such a case the operator $\mathscr{O}_n$ has an explicit form.

Recall that

$$\|\mathscr{O}_n^{-1/2}(f)\|^2_{L^2([-\pi,\pi]^d)} = \lambda_{(\mathrm{sob})}\|f\|^2_{H^s_{\mathrm{per}}} + \lambda_{(\mathrm{pde})}\|\mathscr{D}(f)\|^2_{L^2([-\pi,\pi]^d)} =: \|f\|^2_{\mathrm{RKHS}}$$

- Denote FS the Fourier series operator.
  $\forall$ frequency $k \in \mathbb{Z}^d$, $\quad FS\left[G_n^{-1/2}(f)\right](k) = \sqrt{a_k}\; FS[f](k)$ by parseval.
  where
  $$a_k = \lambda_{(sob)} \sum_{|\alpha| \leq s} \prod_{j=1}^{d} k_j^{2\alpha_j} + \lambda_{(pde)} \left(\sum_{|\alpha| \leq s} p_\alpha \prod_{j=1}^{d} k_j^{\alpha_j} \right)^2$$

- G_n is diagonalizable with eigenfunctions $v_k : x \mapsto \exp(i\langle k, x\rangle)$
  eigenvalues a_k^{-1}.

- Fourier decomposition of f: for all $x \in [-\pi,\pi]^d$
  $$f(x) = \sum_{k \in \mathbb{Z}^d} FS[f](k) \exp(i\langle k, x\rangle)$$
  $$= \sum_{k \in \mathbb{Z}^d} FS\left[G_n^{-1/2}(f)\right](k) \cdot \sqrt{a_k} \cdot \exp(i\langle k, x\rangle)$$

- Set ψ_x such that $FS[\psi_x](k) = (a_k)^{-1/2} \exp(i\langle k, x\rangle)$
  $\psi_x \in H^s_{per}$ since $a_k^{-1} \leq \lambda_{(sob)}^{-1} \left(\sum_{|\alpha| \leq s} \prod_{j=1}^{d} k_j^{2\alpha_j} \right)^{-1} \Rightarrow \sum_{k \in \mathbb{Z}^d} a_k^{-1} < +\infty$

- We have the kernel formulation
  $$f(x) = \left\langle G_n^{-1/2}(f), \psi_x \right\rangle_{L^2([-\pi,\pi]^d)}$$

- $\|G_n^{-1/2}(f)\|_{L^2([-\pi,\pi]^d)} = \|f\|^2_{RKHS}$

↳ The corresponding kernel is

$$K(x,y) = \langle \psi_x, \psi_y \rangle_{L^2([-\pi,\pi]^d)} = \sum_{k \in \mathcal{I}} a_k \, v_k(x) \overline{v_k}(y)$$

In the more general case, this is more technical:

✗ G_n is not diagonal in the Fourier basis

⇝ Need for results of PDE theory.

This result shows that a PIML estimator (and therefore its variants implemented in practice such as PINNs) can be regarded as a kernel estimator.

✗ However computing K is not always straightforward.

- <u>Characterization of the kernel via a weak formulation</u>.

Proposition (Characterization of the kernel)

The kernel K is the unique solution to the following weak formulation, valid for all test functions $\varphi \in H^s_{\mathrm{per}}([-2L, 2L]^d)$: for all $x \in \Omega$,

$$\lambda_{(\mathrm{sob})} \sum_{|\alpha| \leqslant s} \int_{[-2L,2L]^d} \partial^\alpha K(x,\cdot)\, \partial^\alpha \varphi + \lambda_{(\mathrm{pde})} \int_\Omega \mathscr{D}(K(x,\cdot))\, \mathscr{D}(\varphi) = \varphi(x).$$

• Convergence rates

In the previous lecture, we have seen that we have to bound the so-called effective dimension $\mathrm{Tr}\left((\Sigma+\lambda I)^{-1}\Sigma\right)$ for the covariance operator $\Sigma := \mathbb{E}\left[\varphi(X)\otimes\varphi(X)\right]$ or equivalently for the integral operator: $\forall f \in L^2(\Omega, \mathbb{P}_X)$, $\forall x \in \Omega$

$$L_k f(x) = \int_\Omega k(x,y)\, f(y)\, d\mathbb{P}_X(y).$$

$\longrightarrow$ IN THE FORMALISM OF CAPONETTO & VITO.

This is the same thing: $\Sigma : \mathcal{H} \to \mathcal{H}$

Reminder: $(a\otimes b)f = \langle b, f\rangle a$

$$\begin{aligned}
\Sigma f(x) &= \langle k(x,\cdot), \Sigma f\rangle = \langle \varphi(x), \Sigma f\rangle \\
&= \langle \varphi(x), \mathbb{E}[\varphi(X)\otimes\varphi(X)] f\rangle \\
&= \left\langle \varphi(x), \left[\int \varphi(y)\otimes\varphi(y)\, d\mathbb{P}_X(y)\right] f\right\rangle \\
&= \int_\mathcal{X} \langle \varphi(x), (\varphi(y)\otimes\varphi(y))(f)\rangle\, d\mathbb{P}_X(y) \\
&= \int_\mathcal{X} \langle \varphi(x), \langle \varphi(y), f\rangle \varphi(y)\rangle\, d\mathbb{P}_X(y) \\
&= \int_\mathcal{X} \langle \varphi(x), \varphi(y)\rangle \langle \varphi(y), f\rangle\, d\mathbb{P}_X(y) \\
&= \int_\mathcal{X} k(x,y)\, f(y)\, d\mathbb{P}_X(y). \\
&= L_k f(x).
\end{aligned}$$

(!)

(!!) Measurability crisis : what follows is just about handling a panic attack

$X : (\Omega, \mathcal{F}, \mathbb{P}_X) \longrightarrow (\mathcal{H}, \mathcal{B}(\langle \cdot, \cdot \rangle_{\mathcal{H}}))$

$\sigma(\text{"}B_{\|\cdot\|_{\mathcal{H}}})$

measurable random variable

$\mathcal{H}$ Hilbert separable $\Rightarrow \exists$ Hilbertian basis $(e_k)_k$

$\forall f \in \mathcal{H} \quad f = \sum_k \langle f, e_k \rangle e_k$

$\phi : \Omega \to \mathcal{H}$ $\mathcal{C}^0$ (by RKHS property, point evaluation is continuous)

Then $\phi(X)$ measurable.

$Y : \Omega \to \mathcal{H}$ v.a. We want to define $\mathbb{E}(Y)$

Hyp: $\|Y\|_{\mathcal{H}} \leq \alpha$ a.s. $\Rightarrow |\langle Y, e_k \rangle| \leq \alpha$

$\Rightarrow \Omega \to \mathbb{R}$ is a random var.

$\omega \mapsto \langle Y, e_k \rangle(\omega)$

and $\boxed{\mathbb{E}[\langle Y, e_k \rangle] \text{ exists}}$

Def: $\mathbb{E}[Y]$ such that $\boxed{\langle \mathbb{E}[Y], e_k \rangle = \mathbb{E}[\langle Y, e_k \rangle]}$. coordinate-wise OK.

Now: $\forall f \in \mathcal{H} \quad \mathbb{E}[\langle Y, f \rangle] \overset{?}{=} \langle f, \mathbb{E}[Y] \rangle$.

Yes since $\mathbb{E}\langle Y, f \rangle = \mathbb{E}\left[\sum_k \langle Y, e_k \rangle \langle e_k, f \rangle\right]$

$\langle f, \mathbb{E}[Y] \rangle = \left\langle \sum_k \langle f, e_k \rangle e_k, \mathbb{E}[Y] \right\rangle = \sum_k \langle f, e_k \rangle \langle e_k, \mathbb{E}[Y] \rangle$.

$= \sum_k \langle f, e_k \rangle . \mathbb{E}[\langle Y, e_k \rangle]$.

$$\mathbb{E} \sum |\langle Y, e_k\rangle \langle e_k, f\rangle| \leq \mathbb{E} \sqrt{\sum \langle Y, e_k\rangle^2} \sqrt{\sum \langle e_k, f\rangle^2}$$

$$= \mathbb{E} \sqrt{\|Y\|_{\mathcal{H}}^2} \sqrt{\|f\|_{\mathcal{H}}^2}.$$

$$\leq \mathbb{E}\|Y\|_{\mathcal{H}}^2 . \quad \|f\|_{\mathcal{H}}^2 .$$

$$\leq \alpha^2 . \|f\|_{\mathcal{H}}^2 . \quad < +\infty$$

Fubini : $\mathbb{E} \langle Y, f\rangle = \langle \mathbb{E}[Y], f\rangle$.

$$\Sigma f(x) = \langle \varphi(x), \Sigma f\rangle = \langle \varphi(x), \mathbb{E}[\hat{\Sigma} f]\rangle$$

$$= \mathbb{E}\left[\langle \varphi(x), \hat{\Sigma} f\rangle\right]$$

$$= \int k(x,y) f(y) \, d\mathbb{P}_X(y) .$$

end of the panic attack.

Formalism of [Caponnetto & Vito 2007]

Convergence rate of the PIML kernel method

- Integral operator $L_K : L^2(\Omega, \mathbb{P}_X) \to L^2(\Omega, \mathbb{P}_X)$, defined by

$$\forall f \in L^2(\Omega, \mathbb{P}_X), \forall x \in \Omega, \quad L_K f(x) = \int_\Omega K(x,y) f(y) d\mathbb{P}_X(y)$$

- Effective dimension $\mathscr{N}(\lambda_{(\mathrm{sob})}, \lambda_{(\mathrm{pde})}) = \mathrm{tr}(L_K(\mathrm{Id} + L_K)^{-1})$

Theorem (Convergence rate)

Assume that $f^\star \in H^s(\Omega)$, $s > d/2$, $\frac{d\mathbb{P}_X}{dx} \leqslant \kappa$, and the noise ε is (M,σ)-sub-Gamma. Then, for all n large enough,

$$\mathbb{E} \int_\Omega |\hat{f}_n - f^\star|^2 d\mathbb{P}_X \lesssim \log^2(n)\Big(\lambda_{(\mathrm{sob})}\|f^\star\|^2_{H^s(\Omega)} + \lambda_{(\mathrm{pde})}\|\mathscr{D}(f^\star)\|^2_{L^2(\Omega)} + \frac{M^2}{n^2\lambda_{(\mathrm{sob})}} + \frac{\sigma^2 \mathscr{N}(\lambda_{(\mathrm{sob})}, \lambda_{(\mathrm{pde})})}{n}\Big).$$

⚠ Finding the eigenvalues of L_K is not an easy task.

Definition: [projection on Ω]

$C \equiv$ operator on $L^2([-2L, 2L]^d)$ defined by $Cf = f \mathbb{1}_\Omega$

C is a projector ($C^2 = C$) and self-adjoint:

$$\langle f, C(g)\rangle_{L^2([-2L,2L]^d)} = \int_{[-2L,2L]^d} fg \mathbb{1}_\Omega = \langle C(f), g\rangle_{L^2}$$

Thm: $K \equiv$ PIML kernel.

Hyp: $\frac{d\mathbb{P}_X}{dx} \leq \kappa$ (bounded density w.r.t Lebesgue measure)

$$a_m(L_K) \leq \kappa\, a_m(C\mathcal{O}_n C)$$

eigenvalues of L_K eigenvalues of $C\mathcal{O}_n C$

- Effective dimension:

$$\mathcal{N}(\lambda_{(sob)}, \lambda_{(pde)}) = \mathrm{Tr}\left(L_K (I + L_K)^{-1}\right)$$

$$= \sum_{m\in\mathbb{N}} \frac{1}{1 + \left(a_m(L_K)\right)^{-1}}$$

eigenvalue of L_K.

$$\leq \sum_{m\in\mathbb{N}} \frac{1}{1 + \left(\kappa\, a_m(\mathrm{Proj}_\Omega \mathcal{O}_n \mathrm{Proj}_\Omega)\right)^{-1}}$$

eigenvalue of $C\mathcal{O}_n C$

- A first crude bound

Prop: $a_m(L_K) \lesssim \lambda_{(sob)}^{-1}\, m^{-2s/d}$

Remark that $\sum_m a_m < \infty$ if $s > d/2$.

One can bound so that the Sobolev part leads (PDE → weighted Sobolev).

Prop: [Caponnetto & Vito 2007, prop 3]

If $a_m = \mathcal{O}_m(m^{1/6})$, then

$$\sum_{m \in \mathbb{N}} \frac{1}{1 + \lambda a_m} = \mathcal{O}_n(\lambda^{-6})$$

Prop: [Minimum rate]

Set $\lambda_{(sob)} = n^{-2s/2s+d} \log(n)$

$\lambda_{(pde)} \leq n^{-2s/2s+d}$

Then,

$$\mathbb{E} \int_\Omega |\hat{f}_n - f^*|^2 d\mathbb{P}_X = \mathcal{O}_n\left(n^{-2s/2s+d} \log^3(n)\right)$$

✓ The PIML ~~estimator~~ converges at least at the Sobolev minimax rate (up to a log term).

? Can we obtain a better convergence rate? Does the physics help?

• Characterizing the eigenvalues of $C\mathcal{O}_n C$.

Characterization of the eigenvalues

Assumptions:

- $s > d/2$
- $\mathscr{D} = \sum_{|\alpha|\leqslant s} p_\alpha \partial^\alpha$
- a_m, v_m = eigenvalues/eigenfunctions of $\mathrm{Proj}_\Omega\, \mathscr{O}_n\, \mathrm{Proj}_\Omega$
- $w_m \in H^s_{\mathrm{per}}([-2L, 2L]^d)$ extended e.f., s.t. $v_m = a_m^{-1}\mathrm{Proj}_\Omega w_m$

Theorem (Eigenfunction characterization)

For any test function $\varphi \in H^s_{\mathrm{per}}([-2L, 2L]^d)$,

$$\lambda_{(sob)} \sum_{|\alpha|\leqslant s} \int_{[-2L,2L]^d} \partial^\alpha w_m\, \partial^\alpha \varphi + \lambda_{(pde)} \int_\Omega \mathscr{D}(w_m)\, \mathscr{D}(\varphi) = a_m^{-1} \int_\Omega w_m \varphi.$$

In particular, any solution of this weak formulation satisfies:

(i) $\forall x \in \Omega$, $\lambda_{(sob)} \sum_{|\alpha|\leqslant s} (-1)^{|\alpha|} \partial^{2\alpha} w_m(x) + \lambda_{(pde)} \mathscr{D}^* \mathscr{D} w_m(x) = a_m^{-1} w_m(x)$,

(ii) $\forall x \in [-2L, 2L]^d \backslash \bar{\Omega}$, $\sum_{|\alpha|\leqslant s} (-1)^{|\alpha|} \partial^{2\alpha} w_m(x) = 0$.

☞ This weak PDE should be solved in a case-by-case study

• Choice of Sobolev regularization is unimportant

thm: Assume that $s > d/2$. Consider the 3 estimators

$$\hat{f}_n^{(1)} = \underset{f \in H^s(\Omega)}{\mathrm{argmin}} \sum_{i=1}^n |f(X_i) - Y_i|^2 + \lambda_{(sob)} \|f\|^2_{H^s(\Omega)} + \lambda_{(pde)} \|\mathscr{D}(f)\|^2_{L^2(\Omega)}$$

$$\hat{f}_n^{(2)} = \underset{f \in H^s_{per}([-2L,2L]^d)}{\mathrm{argmin}} \sum |f(X_i) - Y_i|^2 + \lambda_{(sob)} \|f\|^2_{H^s_{per}([-2L,2L]^d)} + \lambda_{(pde)} \|\mathscr{D}(f)\|^2_{L^2(\Omega)}$$

$$\hat{f}_n^{(3)} = \underset{f \in H^s(\Omega)}{\mathrm{argmin}} \sum_{i=1}^n |f(X_i) - Y_i|^2 + \lambda_{(sob)} \|f\|^2 + \lambda_{(pde)} \|\mathscr{D}(f)\|^2_{L^2(\Omega)}$$

any equivalent Sobolev norm.

$\hat{f}_n^{(1)}$, $\hat{f}_n^{(2)}$ and $\hat{f}_n^{(3)}$ share equivalent effective dimension $N(\lambda_{(sob)}, \lambda_{(pde)})$
the same UB on the CV rate

The proof relies on the following result:

If $C_1 \left(\lambda_{(sob)} \|f\|^2_{H^s(\Omega)} + \lambda_{(pde)} \|\mathcal{D}(f)\|^2_{L^2(\Omega)}\right) \leq \langle f, f \rangle_n \leq C_2 \left(\lambda_{(sob)} \|f\|^2_{H^s(\Omega)} + \lambda_{(pde)} \|\mathcal{D}(f)\|^2_{L^2(\Omega)}\right)$

Then kernels associated with $\langle \cdot, \cdot \rangle_n$ on $H^s(\Omega)$ have the same CV rate, as the one associated with $\lambda_{(sob)} \|f\|^2_{H^s_{per}([-2L,2L]^d)} + \lambda_{(pde)} \|\mathcal{D}(f)\|^2_{L^2(\Omega)}$.

↳ Need to consider the extension E from $H^s(\Omega)$ to $H^s_{per}([-2L,2L]^d)$ with minimal $H^s_{per}([-2L,2L]^d)$-norm.

↳ Show the inner product equivalence.

- <u>Application : speed-up effect of the physical penalty</u>

Toy setting :
$d = 1$
$\Omega = [-L, L]$
$f^* \in H^1(\Omega)$ $\quad s = 1$
$\mathcal{D} = d/dx$

$\mathcal{D}(f^*) \simeq 0$ means f^* is approximately constant.

$$\hat{f}_n = \underset{f \in H^1_{per}([-2L,2L]^d)}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} |Y_i - f(X_i)|^2 + \lambda_{(sob)} \|f\|^2_{H^1_{per}([-2L,2L]^d)} + \lambda_{(pde)} \|\mathcal{D}(f)\|^2_{L^2(\Omega)}$$

① In this case, the kernel can be analytically computed

Explicit one-dimensional kernel

Let $\gamma_n = \sqrt{\frac{\lambda_{(sob)}}{\lambda_{(sob)}+\lambda_{(pde)}}}$. Then for all $x, y \in [-L, L]$,

$$K(x,y) = \frac{\gamma_n}{2\lambda_n \sinh(2\gamma_n L)}\Big((\cosh(2\gamma_n L) + \cosh(2\gamma_n x))\cosh(\gamma_n(x-y))$$
$$+ ((1 - 2\times 1_{x>y})\sinh(2\gamma_n L) - \sinh(2\gamma_n x))\sinh(\gamma_n(x-y))\Big).$$

② Characterizing the eigenvalues a_m of CO_nC:

To do so, we have to solve the weak formulation

$$\forall \phi \in H^1_{per}([-2L,2L]),\ \lambda_{(sob)} \int_{[-2L,2L]^d} w_m \phi + (\lambda_{(sob)} + \lambda_{(pde)}) \int_\Omega \frac{d}{dx} w_m \frac{d}{dx}\phi = a_m^{-1} \int_\Omega w_m \phi$$

ⓐ Show that $\hat{O}_n = CO_nC$ is symmetric, i.e., $\tilde{O}_n(f)(-x) = \tilde{O}_n(f(-\cdot))(x)$

ⓑ PDE system: the function $w_m \in C^\infty([-L,L])$ satisfy

(i) $\forall x \in \Omega \quad \lambda_{(sob)}\left(1 - \frac{d^2}{dx^2}\right) w_m(x) - \lambda_{(pde)} \frac{d^2}{dx^2} w_m(x) = a_m^{-1} w_m(x)$.

Since $a_m^{-1} \geq \lambda_{(sob)}$, the solutions of this ODE are linear combinations of $\cos\sqrt{\frac{a_m^{-1} - \lambda_{(sob)}}{\lambda_{(sob)} + \lambda_{(pde)}}}\,x$ and $\sin\sqrt{\frac{a_m^{-1} - \lambda_{(sob)}}{\lambda_{(sob)} + \lambda_{(pde)}}}\,x$

(ii) $\forall x \in [-2L, 2L]^d \setminus \bar{\Omega} \qquad \left(1 - \frac{d^2}{dx^2}\right) w_m(x) = 0$

The solutions of this ODE are linear combination of cosh and sinh.

And the $\mathcal{C}^\infty$ $4L$-periodic junction condition at $-2L$

$\forall\ -2L \leqslant x \leqslant -L \qquad w_m(x) = A\cosh(x+2L) + B\sinh(x+2L)$

$\forall\ \quad L \leqslant x \leqslant 2L \qquad w_m(x) = A\cosh(x-2L) + B\sinh(x-2L)$

ⓒ Symmetric eigenfunctions

$$w_m(x) = \begin{cases} A\cosh(x+2L) & \text{if} \quad -2L \leqslant x \leqslant -L \\ C\cos\left(\sqrt{\dfrac{a_m^{-1} - \lambda_{(sob)}}{\lambda_{(sob)} + \lambda_{(pde)}}}\, x\right) & \text{if} \quad -L \leqslant x \leqslant L \\ A\cosh(x-2L) & \text{if} \quad L \leqslant x \leqslant 2L \end{cases}$$

System at $x = -L$ $\qquad A\cosh(L) = C\cos\left(\sqrt{\dfrac{\left(a_m^{sym}\right)^{-1} - \lambda_{(sob)}}{\lambda_{(sob)} + \lambda_{(pde)}}}\, L\right)$

+ condition on the derivative

$$\left(\lambda_{(sob)} + \lambda_{(pde)}\right) \lim_{\substack{x \to -L \\ x > -L}} \frac{d}{dx} w_m(x) = \lambda_{(sob)} \lim_{\substack{x \to -L \\ x < -L}} \frac{d}{dx} w_m(x)$$

Deduce that

$$\lambda_{(sob)} + \left(\lambda_{(sob)} + \lambda_{(pde)}\right)\left(m - \frac{1}{2}\right)^2 \frac{\pi^2}{L^2} \leqslant \left(a_m^{sym}\right)^{-1} \leqslant \lambda_{(sob)} + \left(\lambda_{(sob)} + \lambda_{(pde)}\right)\left(m + \frac{1}{2}\right)^2 \frac{\pi^2}{L^2}.$$

ⓓ Antisymmetric eigenfunctions ...

$$\lambda_{(sob)} + \left(\lambda_{(sob)} + \lambda_{(pde)}\right)\left(m - \frac{1}{2}\right)^2 \frac{\pi^2}{L^2} \leqslant \left(a_m^{anti}\right)^{-1} \leqslant \lambda_{(sob)} + \left(\lambda_{(sob)} + \lambda_{(pde)}\right)\left(m + \frac{1}{2}\right)^2 \frac{\pi^2}{L^2}$$

$\left(\lambda_{(sob)} + \lambda_{(pde)}\right)\left(m - \frac{1}{2}\right)^2 \frac{\pi^2}{L^2} \leqslant$

② Conclusion : putting the bounds obtained for a_m^{sym} and a_m^{anti} together

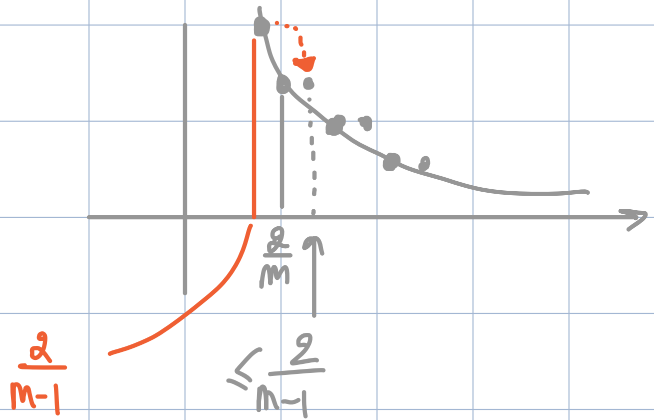
We have 2 decreasing sequences to combine.

$m \leftarrow m/2$

$(m/2-1)^2 = \frac{(m-2)^2}{4}$

et λ_n à mettre ac la rcd.

$a_m = 1/m$ $\quad b_m = 1/m$ $\quad c_m =$ concatenation of a_m and b_m, decreasing sequences.



$$c_m = \begin{cases} a_{m/2} & m \text{ even.} \\ b_{(m-1)/2} & m \text{ odd.} \end{cases}$$

Prop

$$\frac{4L^2}{(\lambda_{(sob)}+\lambda_{(pde)})(m+4)^2\pi^2} \leq a_m \leq \frac{4L^2}{(\lambda_{(sob)}+\lambda_{(pde)})(m-2)^2\pi^2}$$

Summary:

$$a_m = \Theta\left(\frac{4L^2}{(\lambda_{(sob)}+\lambda_{(pde)})\, m^2\pi^2}\right)$$

thm: (Kernel speed-up)

Theorem (Kernel speed-up)

Let $\lambda_{(sob)} = n^{-1}\log(n)$ *and*

$$\lambda_{(pde)} = \begin{cases} n^{-2/3}/\|\mathscr{D}(f^\star)\|_{L^2(\Omega)} & \text{if} \quad \|\mathscr{D}(f^\star)\|_{L^2(\Omega)} \neq 0 \\ 1/\log(n) & \text{if} \quad \|\mathscr{D}(f^\star)\|_{L^2(\Omega)} = 0. \end{cases}$$

Then

$$\mathbb{E}\int_{[-L,L]} |\hat{f}_n - f^\star|^2 d\mathbb{P}_X = \|\mathscr{D}(f^\star)\|_{L^2(\Omega)}\, \mathcal{O}_n\big(n^{-2/3}\log^3(n)\big)$$
$$+ \big(\|f^\star\|^2_{H^s(\Omega)} + \underbrace{\sigma^2 + M^2}_{\textit{noise param}}\big)\, \mathcal{O}_n\big(n^{-1}\log^3(n)\big).$$

- ✓ When $\|\mathscr{D}(f^\star)\|_{L^2(\Omega)} = 0$ ($f^\star$ is constant), the PIML method recovers the parametric convergence rate of n^{-1}
- ✓ When $\|\mathscr{D}(f^\star)\|_{L^2(\Omega)} > 0$, we recover the Sobolev minimax convergence rate in $H^1(\Omega)$ of $n^{-2/3}$

Incorporating the physics in the learning process helps learning!