Local approximation of operators

H. N. Mhaskar
Claremont Graduate University, Claremont, CA
hrushikesh.mhaskar@cgu.edu

ICERM Workshop:Mathematical and Scientific Machine Learning
June 7, 2023

Problem statement

$\mathfrak{X}, \mathfrak{Y}$: Metric spaces
$K_{\mathfrak{X}}$: Compact subset of $\mathfrak{X}$
$\mathcal{F} : K_{\mathfrak{X}} \to \mathfrak{Y}$: continuous function, with some smoothness to be explored.

Goal:

Given a finite amount of information about $\mathcal{F}$,

- find efficient methods to approximate $\mathcal{F}$
- estimate the degree of approximation.

An existence theorem[1]

Let σ be an activation function admitting universal approximation, X Banach space, $K_1 \subset X$, $K_2 \subset \mathbb{R}^q$ compact, $\mathfrak{F} : C(K_1) \to C(K_2)$. Then, for $\epsilon > 0, f \in C(K_1)$, there exists a network of the form

$$\mathfrak{G}(f)(y) = \sum_{k=1}^{N} \underbrace{\sum_{j=1}^{M} c_{j,k}\sigma\left(\sum_{\ell=1}^{M} \xi_{j,k,\ell} f(x_\ell) + \theta_{j,k}\right)}_{branch} \underbrace{\sigma\left(\omega_k \cdot y + \zeta_k\right)}_{trunk}$$

such that

$$|\mathfrak{F}(f)(y) - \mathfrak{G}(f)(y)| \le \epsilon, \quad y \in K_2.$$

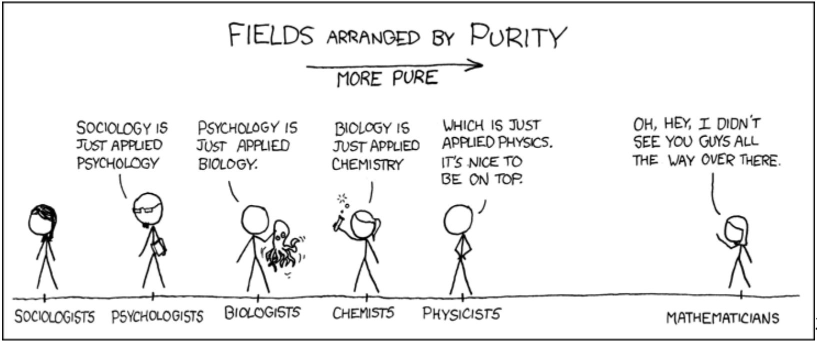
[1] Chen, Chen, 1995

Observations

- This is an existence theorem only.
- Requires values f as input
- Does the whole approximation in one stroke
- Might not be always a good idea.
  - Degree of approximation with constructions were also known[2]

[2] Mhaskar, Hahm, 1996

Problem reduction

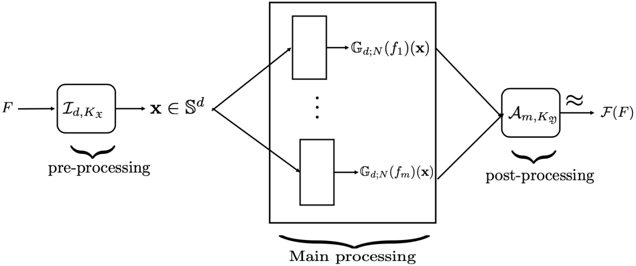


[3]

They don’t see me either!

[3] https://imgs.xkcd.com/comics/purity.png

Problem reduction



Prima facie complexity: *dNm*.

Problem reduction

$\mathcal{F} : \mathfrak{X} \to \mathfrak{Y}$, $\mathcal{F}(K_{\mathfrak{X}}) \subset K_{\mathfrak{Y}}$.

$$\mathcal{I}_{d,\mathfrak{X}} : \mathfrak{X} \to \mathbb{R}^d, \quad \mathcal{A}_{d,\mathfrak{X}} : \mathbb{R}^d \to \mathfrak{X},$$

$$\max_{F \in K_{\mathfrak{X}}} \rho_{\mathfrak{X}} \left(F, \mathcal{A}_{d,\mathfrak{X}}(\mathcal{I}_{d,\mathfrak{X}}(F))\right) \lesssim \mathrm{width}_d(K_{\mathfrak{X}}, \mathfrak{X}). \quad \text{Hope!}$$

$$\mathcal{I}_{m,\mathfrak{Y}} : \mathfrak{Y} \to \mathbb{R}^m, \quad \mathcal{A}_{m,\mathfrak{Y}} : \mathbb{R}^m \to \mathfrak{Y},$$

$$\max_{F \in K_{\mathfrak{X}}} \rho_{\mathfrak{Y}} \left(\mathcal{F}(F), \mathcal{A}_{m,\mathfrak{Y}}(\mathcal{I}_{m,\mathfrak{Y}}(\mathcal{F}(F)))\right) \lesssim \mathrm{width}_m(K_{\mathfrak{Y}}, \mathfrak{Y}) \quad \text{Hope again!}$$

Approximate the function $\mathbf{f} : \mathbb{R}^d \to \mathbb{R}^m$, $\mathbf{f}(\mathcal{I}_{d,\mathfrak{X}}(F)) = \mathcal{I}_{m,\mathfrak{Y}}(\mathcal{F}(F))$.

1. Example[3]

[3] Mhaskar, Prestin 2000, 2005, Mhaskar, Nevai, Shvarts, 2013, Mhaskar, 2020

Frame operator

Let $h : \mathbb{R} \to [0, 1]$ be infinitely often differentiable even function, nonincreasing on $[0, \infty)$, $h(t) = 1$ if $|t| \le 1/2$, $h(t) = 0$ if $|t| \ge 1$. Let

$$g(t) = \sqrt{h(t) - h(2t)},$$

$$\Psi_j^*(x) = \sum_k g\left(\frac{|k|}{2^j}\right) \exp(ikx) \in \mathbb{H}_{2^j-1}, j \ge 1,$$

$$\Psi_0^*(x) = 1.$$

$$\tau_j^*(f, x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \Psi_j^*(x - t) dt.$$

$$x_{j,k} = \frac{2\pi k}{2^{j+1}}, \quad c_{j,k}(f) = \tau_j^*(f, x_{j,k}) = \sum_\ell g\left(\frac{|\ell|}{2^j}\right) \hat{f}(\ell) \exp(i\ell x_{j,k}).$$

Wavelet-like expansion

Let $f \in C^*$. Then for $x \in [-\pi, \pi]$,

$$f(x) = \sum_{j=0}^{\infty} 2^{-j-1} \sum_{k=0}^{2^{j+1}-1} c_{j,k}(f) \Psi_j^*(x - x_{j,k})$$

where the series converges uniformly. We have

$$\sum_{j=0}^{\infty} 2^{-j-1} \sum_{k=0}^{2^{j+1}-1} |c_{j,k}(f)|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(t)|^2 dt.$$

Characterization of smoothness

$\gamma = r + \alpha$, $r \geq 0$ integer, $0 < \alpha \leq 1$.

$f \in W_\gamma$ if f has r continuous derivatives and

$$\sup_{x \in [-\pi,\pi]} |f(x+h) + f(x-h) - 2f(x)| = \mathcal{O}(|h|^\alpha).$$

$f \in W_\gamma(x_0)$ if there is $I \ni x_0$ such that for every $\phi \in C^\infty(\mathbb{T})$, supported on I, $\phi f \in W_\gamma$.

$f \in W_\gamma$ if and only if

$$\max_{0 \le k \le 2^{j+1}} |c_{j,k}(f)| = \mathcal{O}(2^{-j\gamma}).$$

Let $x_0 \in [-\pi, \pi]$, $\gamma > 0$, $f \in C^*$. We have $f \in W_\gamma(x_0)$ if and only if there is a nondegenerate interval $I \ni x_0$ such that

$$\max_{x_{j,k} \in I} |c_{j,k}(f)| = \mathcal{O}(2^{-j\gamma}).$$

Remark: Similar theorems are known for general manifolds [4].

[4] Mh. 2020

Encoder/decoder

$\mathfrak{F} : C(\mathbb{T}) \to C(\mathbb{T})$, $d = 2^n$, $m = 2^L$,

- Information on $F \in C(\mathbb{T})$: $\{\hat{F}(\ell)\}_{|\ell|<2^n}$

$$\mathcal{I}_{d,C(\mathbb{T})} = \{c_{j,k}(F)\}_{k=0,\cdots,2^{j+1}-1,\ j=0,\cdots,2^n}$$

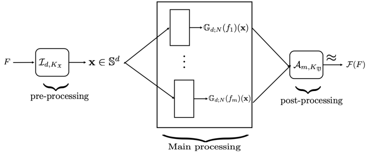
- Approximate

$$\{c_{j,k}(\mathfrak{F}(F))\}_{k=0,\cdots,2^{j+1}-1,\ j=0,\cdots,2^L}$$

- Reconstruct

$$\sum_{j=0}^{L} 2^{-j-1} \sum_{k=0}^{2^{j+1}-1} c_{j,k}(\mathfrak{F}(F))\Psi_j^*(x - x_{j,k}), \quad x_{j,k} = \frac{2\pi k}{2^{j+1}}$$

Problems in general theory



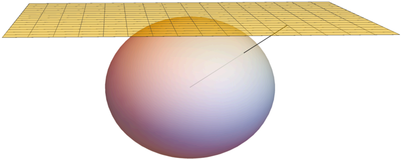
d, m, and the complexity of approximation need all to be large.

Solutions

- Assume extra smoothness on $\mathbf{f}$. (Caution: the dependence on d).
- Local approximation
    - Use only values in a small neighborhood of F to approximate $\mathcal{F}(F)$ (Distributed learning)
    - The approximation should adjust automatically to the local smoothness of $\mathbf{f}$.

Conversion to the sphere

$$\mathbb{S}^d = \{\mathbf{x} \in \mathbb{R}^{d+1} : |\mathbf{x}|_{d+1} = 1\}, \qquad \mathbb{S}^d_+ = \{\mathbf{x} \in \mathbb{S}^d : x_{d+1} > 0\}.$$



Coordinate chart for $\mathbb{S}^d_+$:

$$\pi^*(x_1, \cdots, x_d) = (x_1, \cdots, x_d, 1)\left(1 + |\mathbf{x}|^2\right)^{-1/2}.$$

Focus on approximation of $f : \mathbb{S}^d \to \mathbb{R}$.

Ingredients

- Jacobi and spherical polynomials
- Definition of smoothness
- Quadrature formula
- Kernels

Notation

μ_d^*=volume measure on $\mathbb{S}^d$, $\mu_d^*(\mathbb{S}^d)=1$, ω_d=volume of $\mathbb{S}^d$.
Π_n^d=set of spherical polynomials of degree $< n$ (restrictions to $\mathbb{S}^d$ of $(d+1)$-variate polynomials of total degree $< n$).
Jacobi polynomials $p_\ell^{(\alpha,\beta)}$ univariate polynomial of degree $=\ell$,

$$\int_{-1}^{1} p_\ell^{(\alpha,\beta)}(x)p_j^{(\alpha,\beta)}(x)(1-x)^\alpha(1+x)^\beta dx=\delta_{\ell,j}.$$

$$K_{d;n}(x)=\frac{2\sqrt{\pi}\Gamma((d+2)/2)}{\Gamma((d+1)/2)(2n+d-2)}p_{n-1}^{(d/2,d/2-1)}(1)p_{n-1}^{(d/2,d/2-1)}(x)$$

$$P(\mathbf{x})=\int_{\mathbb{S}^d} P(\mathbf{y})K_{d;n}(\mathbf{x}\cdot\mathbf{y})d\mu_d^*(\mathbf{y}),\qquad P\in\Pi_n^d.$$

Smoothness classes

Lipschitz condition:

$$|f(\mathbf{y}) - f(\mathbf{x})| \leq c|\mathbf{y} - \mathbf{x}|_{d+1};$$

i.e.,

$$\max_{\mathbf{y}\in\mathbb{S}^d} \frac{|f(\mathbf{y}) - f(\mathbf{x})|}{|\mathbf{y} - \mathbf{x}|_{d+1}} < \infty.$$

Treating $f(\mathbf{x}) \in \Pi_1^d$ (constant functions),

$$\min_{P\in\Pi_1^d} \max_{\mathbf{y}\in\mathbb{S}^d} \frac{|f(\mathbf{y}) - P(\mathbf{y})|}{|\mathbf{y} - \mathbf{x}|_{d+1}} < \infty.$$

Smoothness classes

Let $f \in C(\mathbb{S}^d)$, $r > 0$ and $\mathbf{x} \in \mathbb{S}^d$. The function f is said to be r-smooth at $\mathbf{x}$ if there exists $\delta = \delta(d; f, \mathbf{x}) > 0$ such that

$$\|f\|_{d;r,\mathbf{x}} := \|f\|_\infty + \min_{P \in \Pi_r^d} \max_{\mathbf{y} \in \mathbb{B}(\mathbf{x},\delta)} \frac{|f(\mathbf{y}) - P(\mathbf{y})|}{|\mathbf{x} - \mathbf{y}|_{d+1}^r} < \infty.$$

$$\|f\|_{d;r} = \sup_{\mathbf{x} \in \mathbb{S}^d} \|f\|_{d;r,\mathbf{x}} < \infty.$$

$$W_{d;r,\mathbf{x}} = \{f \in C(\mathbb{S}^d) : \|f\|_{d;r,\mathbf{x}} < \infty\}, \quad W_{d;r} = \{f \in C(\mathbb{S}^d) : \|f\|_{d;r} < \infty\}.$$

Smoothness classes

Let $f : [-1,1] \to \mathbb{C}$. Then f has an analytic extension to $\{z \in \mathbb{C} : |z + \sqrt{z^2 - 1}| < e^{\rho}\}$ if and only if

$$\limsup_{n\to\infty} \left\{ \min_{P \in \Pi_n} \|f - P\|_{\infty,[-1,1]} \right\}^{1/n} = e^{-\rho} < 1.$$

Smoothness classes

Let $f \in C(\mathbb{S}^d)$, $\mathbf{x} \in \mathbb{S}^d$, $\rho > 0$. The function f is said to be ρ-analytic at x if there is exists $\delta = \delta(d; f, \mathbf{x}) > 0$ such that

$$\|f\|_{A_{d;\rho,\mathbf{x}}} = \|f\|_\infty + \sup_{n \geq 0} \left\{ \exp(\rho n) \min_{P \in \Pi_n^d} \|f - P\|_{\infty, \mathbb{B}(\mathbf{x},\delta)} \right\} < \infty.$$

$$A_{d;\rho,\mathbf{x}} = \{f \in C(\mathbb{S}^d) : \|f\|_{A_{d;\rho,\mathbf{x}}} < \infty.\}$$

$$\|f\|_{A_{d;\rho}} = \|f\|_\infty + \sup_{n > 0} \{\exp(\rho n) E_{d;n}(f)\}$$

$$A_{d;\rho} = \{f \in C(\mathbb{S}^d) : \|f\|_{A_{d;\rho}} < \infty\}.$$

MZ quadrature formula

Let $n \geq 1$. A measure ν on $\mathbb{S}^d$ is called Marcinkiewicz-Zygmund quadrature measure of order n ($\nu \in \mathrm{MZQ}(d;n)$) if

$$\int_{\mathbb{S}^d} P d\nu = \int_{\mathbb{S}^d} P d\mu_d^*, \quad P \in \Pi_n^d,$$

and

$$\int_{\mathbb{S}^d} |P|^2 d|\nu| \leq \|\nu\|_{d;n} \int_{\mathbb{S}^d} |P|^2 d\mu_d^*, \quad P \in \Pi_{n/2}^d$$

MZ quadrature formula

Let $\mathcal{C} \subset \mathbb{S}^d$. There exists $C = C(d)$ such that if

$$\delta(\mathcal{C}) = \max_{\mathbf{x}\in\mathbb{S}^d} \min_{\mathbf{y}\in\mathcal{C}} \rho(\mathbf{x}, \mathbf{y}) \le C/n,$$

then there exists[5] a $\nu \in \mathrm{MZQ}(d; n)$ supported on $\mathcal{C}$.

[5] Mhaskar, Narcowich, Ward, 2001, Filbir, Mhaskar, 2011

Let $n \geq 1$. There exist positive numbers w_k, and points $\mathbf{y}_k$, $k = 1, \cdots, \dim(\Pi_n^d)$, such that

$$\sum_{k=1}^{\dim(\Pi_n^d)} w_k P(\mathbf{y}_k) = \int_{\mathbb{S}^d} P(\mathbf{y}) d\mu_d^*(\mathbf{y}), \quad P \in \Pi_n^d.$$

Remark. If ν_n^* is the measure associating the mass w_k with $\mathbf{y}_k$, then $\nu_n^* \in \mathrm{MZQ}(d; n)$.

Kernel

$$\Phi_{d;n,r}(x) = K_{d;(d+2)n}(x)\frac{p_{dn}^{(d/2+r,d/2-2)}(x)}{p_{dn}^{(d/2+r,d/2-2)}(1)}\left(\frac{1+x}{2}\right)^n.$$

$$\sigma_{d;n,r}(\nu,f)(\mathbf{x}) = \int_{\mathbb{S}^d} f(\mathbf{y})\Phi_{d;n,r}(\mathbf{x}\cdot\mathbf{y})d\nu(\mathbf{y})$$

Remark:

- The measure ν depends only on the locations at which f is sampled, not f itself. It is a pre-computation.
- The construction is universal approximation; defined for all f, without requiring prior assumptions on smoothness of f.

Kernel

$$\Phi_{d;n,r}(x) = K_{d;(d+2)n}(x)\frac{p_{dn}^{(d/2+r,d/2-2)}(x)}{p_{dn}^{(d/2+r,d/2-2)}(1)}\left(\frac{1+x}{2}\right)^n.$$

$$\sigma_{d;n,r}(\nu,f)(\mathbf{x}) = \int_{\mathbb{S}^d} f(\mathbf{y})\Phi_{d;n,r}(\mathbf{x}\cdot\mathbf{y})d\nu(\mathbf{y})$$

$$E_{d;n}(f) = \min_{P\in\Pi_n^d} \|f - P\|_\infty.$$

Theorem

If $\nu \in \mathrm{MZQ}(d; 2(d+2)n)$, then for $f \in C(\mathbb{S}^d)$,

$$E_{d;2(d+2)n}(f) \leq \|f - \sigma_{d;n,r}(\nu;f)\|_\infty \lesssim d^{1/6}|||\nu|||_{d;2(d+2)n}E_{d;n}(f).$$

Let $d \geq 4$, $\mathbf{x} \in \mathbb{S}^d$, $r = r(\mathbf{x}) > 0$,and $f \in W_{d;r,\mathbf{x}}$. Let $nd \geq (d + r + 1)^2$, $\nu \in \mathrm{MZQ}(d; 2(d+2)n)$. If n is large enough so that

$$\delta_n = \sqrt{\frac{16r \log n}{n}} \leq \delta(d; f, \mathbf{x}),$$

then

$$|f(\mathbf{x}) - \sigma_{d;n,r}(\nu; f)(\mathbf{x})| \lesssim \frac{d^{1/6} \|f\|_{W_{d;r,\mathbf{x}}} |\!|\!|\nu|\!|\!|_{d;2(d+2)n}}{\dim(\Pi^d_{2(d+2)n})^{r/d}}.$$

Moreover,

$$\left| f(\mathbf{x}) - \int_{\mathbb{B}(x,\delta_n)} \Phi_{d;n,r}(\mathbf{x} \cdot \mathbf{y}) f(\mathbf{y}) d\nu(\mathbf{y}) \right| \lesssim \frac{d^{1/6} \|f\|_{W_{d;r,\mathbf{x}}} |\!|\!|\nu|\!|\!|_{d;2(d+2)n}}{\dim(\Pi^d_{2(d+2)n})^{r/d}}.$$

Local approximation

Let $d \geq 4$, $r \geq 0$, $nd \geq (d+r+1)^2$, $\nu \in \mathrm{MZQ}(d; 2(d+2)n)$. If $\mathbf{x} \in \mathbb{S}^d$, $f \in A_{d;\rho,\mathbf{x}}$, and $\delta = \delta(d; f, \mathbf{x})$ (as in definition of $A_{d;\rho,\mathbf{x}}$). Then with

$$\Delta = \min(\rho, \delta^2/4 - 2\log(4/\delta)),$$

$$|f(\mathbf{x}) - \sigma_{d;n,r}(\nu; f)(\mathbf{x})| \lesssim d^{1/6} \exp(-n\Delta) \|f\|_{A_{d;\rho,\mathbf{x}}} |\!|\!|\nu|\!|\!|_{2(d+2)n},$$

and

$$\left| f(\mathbf{x}) - \int_{\mathbb{B}(\mathbf{x},\delta(d;f,\mathbf{x}))} \Phi_{d;n,r}(\mathbf{x}\cdot\mathbf{y}) f(\mathbf{y}) d\nu(\mathbf{y}) \right| \lesssim d^{1/6} \exp(-n\Delta) \|f\|_{A_{d;\rho,\mathbf{x}}} |\!|\!|\nu|\!|\!|_{2(d+2)n}.$$

From polynomials to networks

$\phi : [-1, 1] \to \mathbb{R}$, $\phi(t) \sim \sum_{\ell=0}^{\infty} \hat{\phi}(\ell) p_\ell^{(d/2-1,d/2-1)}(1) p_\ell^{(d/2-1,d/2-1)}(t)$.

Any $P \in \Pi_n^d$ (with degree n) can be exptessed in the form $P(\mathbf{x}) = \sum_{\ell=0}^{n} P_\ell(\mathbf{x})$, where P_ℓ is homogeneous, harmonic polynomial of total degree ℓ.

Let

$$\mathcal{D}_\phi(P)(\mathbf{x}) = \sum_{\ell=0}^{n} \hat{\phi}(\ell)^{-1} P_\ell(\mathbf{x}),$$

Then

$$P(\mathbf{x}) = \int_{\mathbb{S}^d} \phi(\mathbf{x} \cdot \mathbf{z}) \mathcal{D}_\phi(P)(\mathbf{z}) d\mu_d^*(\mathbf{z}).$$

From polynomials to networks

$$\Phi_{d;n,r}(\mathbf{x}\cdot\mathbf{y}) = \int_{\mathbb{S}^d} \phi(\mathbf{x}\cdot\mathbf{z})\mathcal{D}_\phi(\Phi_{d;n,r}(\circ,\mathbf{y}))(\mathbf{z})d\mu_d^*(\mathbf{z}).$$

Discretization leads to a (pre-fabricated) zonal function network of the form $\sum_k a_k(\mathbf{y})\phi(\mathbf{x}\cdot\mathbf{z}_k)$, where [6]

- $\mathbf{z}_k$ are fixed independent of $\mathbf{x}$,
- a_k are pre-computed functions of y
- The size of the network $\sim \dim(\Pi_n^d)$.
- No training is required.

TINN for PINN?

[6] Mhaskar, Narcowich, Ward, 1999, Mhaskar, 2006, 2010, 2019, 2020, 2020

Thank you.