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

\( \mathcal{X}, \mathcal{Y} \): Metric spaces
\( K_\mathcal{X} \): Compact subset of \( \mathcal{X} \)
\( \mathcal{F} : K_\mathcal{X} \rightarrow \mathcal{Y} \): 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 \( \sigma \) be an activation function admitting universal approximation, \( X \) Banach space, \( K_1 \subset X, K_2 \subset \mathbb{R}^q \) compact, \( \mathcal{F} : C(K_1) \rightarrow C(K_2) \). Then, for \( \epsilon > 0, f \in C(K_1) \), there exists a network of the form

\[
\mathcal{G}(f)(y) = \sum_{k=1}^{N} \sum_{j=1}^{M} c_{j,k} \sigma \left( \sum_{\ell=1}^{M} \xi_{j,k,\ell} f(x_\ell) + \theta_{j,k} \right) \sigma \left( \omega_k \cdot y + \zeta_k \right)
\]

such that

\[
|\mathcal{F}(f)(y) - \mathcal{G}(f)(y)| \leq \epsilon, \quad y \in K_2.
\]

---

\(^1\) 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

They don’t see me either!

3 https://imgs.xkcd.com/comics/purity.png
Problem reduction

Prima facie complexity: $dNm$. 

\[ F \xrightarrow{\mathcal{I}_{d,K_x}} x \in S^d \]

Main processing

pre-processing

\[ g_{d,N}(f_1)(x) \]

\[ \cdots \]

\[ g_{d,N}(f_m)(x) \]

post-processing

\[ \mathcal{A}_{m,K_{2\eta}} \xrightarrow{\sim} F(F) \]
Problem reduction

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

$$I_{d,\mathcal{X}} : \mathcal{X} \to \mathbb{R}^d, \quad A_{d,\mathcal{X}} : \mathbb{R}^d \to \mathcal{X},$$

$$\max_{F \in K_{\mathcal{X}}} \rho_{\mathcal{X}} (F, A_{d,\mathcal{X}}(I_{d,\mathcal{X}}(F))) \lesssim \text{width}_d(K_{\mathcal{X}}, \mathcal{X}). \quad \text{Hope!}$$

$$I_{m,\mathcal{Y}} : \mathcal{Y} \to \mathbb{R}^m, \quad A_{m,\mathcal{Y}} : \mathbb{R}^m \to \mathcal{Y},$$

$$\max_{F \in K_{\mathcal{X}}} \rho_{\mathcal{Y}} (\mathcal{F}(F), A_{m,\mathcal{Y}}(I_{m,\mathcal{Y}}(\mathcal{F}(F)))) \lesssim \text{width}_m(K_{\mathcal{Y}}, \mathcal{Y}) \quad \text{Hope again!}$$

Approximate the function $$f : \mathbb{R}^d \to \mathbb{R}^m, \ f(I_{d,\mathcal{X}}(F)) = I_{m,\mathcal{Y}}(\mathcal{F}(F)).$$
1. Example³

³Mhaskar, Prestin 2000, 2005, Mhaskar, Nevai, Shvarts, 2013, Mhaskar, 2020
Let \( h : \mathbb{R} \to [0, 1] \) be infinitely often differentiable even function, nonincreasing on \([0, \infty)\), \( h(t) = 1 \) if \(|t| \leq 1/2\), \( h(t) = 0 \) if \(|t| \geq 1\). Let

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

\[
\Psi^*_j(x) = \sum_k g \left( \frac{|k|}{2^j} \right) \exp(ikx) \in \mathbb{H}_{2^{j-1}}, j \geq 1,
\]

\[
\Psi^*_0(x) = 1.
\]

\[
\tau^*_j(f, x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \Psi^*_j(x - t) dt.
\]

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

\[ \gamma = r + \alpha, \quad r \geq 0 \text{ integer}, \quad 0 < \alpha \leq 1. \]

\( f \in W_\gamma \) if \( f \) has \( r \) continuous derivatives and

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

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

\( \phi f \in W_\gamma. \)
Characterization of smoothness

\( f \in W_\gamma \) if and only if

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

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 \(^4\).
\[ \mathcal{F} : C(\mathbb{T}) \rightarrow C(\mathbb{T}), \quad d = 2^n, \quad m = 2^L, \]

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

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

- Approximate

\[ \{ c_{j,k}(\mathcal{F}(F)) \}_{k=0, \ldots, 2^{j+1}-1, j=0, \ldots, 2^L} \]

- Reconstruct

\[ \sum_{j=0}^L 2^{-j-1} \sum_{k=0}^{2^{j+1}-1} c_{j,k}(\mathcal{F}(F)) \Psi^*_j(x - x_{j,k}), \quad x_{j,k} = \frac{2\pi k}{2^{j+1}} \]
$d$, $m$, and the complexity of approximation need all to be large.

**Solutions**

- **Assume extra smoothness on** $f$. (*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 $f$. 
Conversion to the sphere

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

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

$\mu_d^*=$volume measure on $S^d$, $\mu_d^*(S^d) = 1$, $\omega_d=$volume of $S^d$.

$\Pi_d^n=$set of spherical polynomials of degree $< n$ (restrictions to $S^d$ of $(d + 1)$-variate polynomials of total degree $< n$).

Jacobi polynomials $p^{(\alpha, \beta)}_\ell$ univariate polynomial of degree $= \ell$,

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

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

$$
P(x) = \int_{S^d} P(y)K_{d;n}(x \cdot y)\, d\mu_d^*(y), \quad P \in \Pi_d^n.
$$
Lipschitz condition:

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

i.e.,

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

Treating \( f(x) \in \Pi_1^d \) (constant functions),

\[ \min_{P \in \Pi_1^d} \max_{y \in \mathbb{S}^d} \frac{|f(y) - P(y)|}{|y - x|_{d+1}} < \infty. \]
Let $f \in C(\mathbb{S}^d)$, $r > 0$ and $x \in \mathbb{S}^d$. The function $f$ is said to be $r$-smooth at $x$ if there exists $\delta = \delta(d; f, x) > 0$ such that

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

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

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

\[
\limsup_{n \to \infty} \left\{ \min_{P \in \Pi_n} \|f - P\|_{\infty, [-1, 1]} \right\}^{1/n} = e^{-\rho} < 1.
\]
Let $f \in C(S^d)$, $x \in S^d$, $\rho > 0$. The function $f$ is said to be $\rho$-analytic at $x$ if there is exists $\delta = \delta(d; f, x) > 0$ such that

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

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

$$
\|f\|_{A_d; \rho} = \|f\|_{\infty} + \sup_{n > 0} \{ \exp(\rho n) E_d; n(f) \}
$$

$$
A_{d; \rho} = \{ f \in C(S^d) : \|f\|_{A_{d; \rho}} < \infty \}.
$$
Let $n \geq 1$. A measure $\nu$ on $S^d$ is called \textit{Marcinkiewicz-Zygmund quadrature measure of order $n$} ($\nu \in \text{MZQ}(d; n)$) if

$$\int_{S^d} P d\nu = \int_{S^d} P d\mu^*_d, \quad P \in \Pi^n_d,$$

and

$$\int_{S^d} |P|^2 d|\nu| \leq \|\nu\|_{d; n} \int_{S^d} |P|^2 d\mu^*_d, \quad P \in \Pi^{n/2}_d.$$
Let $\mathcal{C} \subset \mathbb{S}^d$. There exists $C = C(d)$ such that if

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

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

\(^5\)Mhaskar, Narcowich, Ward, 2001, Filbir, Mhaskar, 2011
Let $n \geq 1$. There exist positive numbers $w_k$, and points $y_k$, $k = 1, \cdots, \dim(\Pi_n^d)$, such that

$$\dim(\Pi_n^d) \sum_{k=1}^{\dim(\Pi_n^d)} w_k P(y_k) = \int_{S^d} P(y) d\mu^*_d(y), \quad P \in \Pi_n^d.$$ 

**Remark.** If $\nu_n^*$ is the measure associating the mass $w_k$ with $y_k$, then $\nu_n^* \in MZQ(d; n)$. 
Kernel

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

\[ \sigma_{d;n,r}(\nu,f)(x) = \int_{\mathbb{S}_d} f(y) \Phi_{d;n,r}(x \cdot y) d\nu(y) \]

Remark:

- The measure \( \nu \) 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

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

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

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

**Theorem**

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

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

then

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

Moreover,

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

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

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

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

$$\left|f(x) - \int_{\mathbb{B}(x, \delta(d; f, x))} \Phi_{d; n, r}(x \cdot y)f(y)d\nu(y)\right|$$

$$\lesssim d^{1/6} \exp(-n\Delta) \|f\|_{A_{d; \rho, x}} \|\nu\|_{2(d+2)n}.$$
\( \phi : [-1, 1] \to \mathbb{R}, \phi(t) \sim \sum_{\ell=0}^{\infty} \hat{\phi}(\ell) p_{\ell}^{(d/2-1,d/2-1)}(1)p_{\ell}^{(d/2-1,d/2-1)}(t) \).

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

Let
\[
D_\phi(P)(x) = \sum_{\ell=0}^{n} \hat{\phi}(\ell)^{-1} P_\ell(x),
\]

Then
\[
P(x) = \int_{S^d} \phi(x \cdot z) D_\phi(P)(z) d\mu^*_d(z).
\]
From polynomials to networks

\[ \Phi_{d;n,r}(x \cdot y) = \int_{S^d} \phi(x \cdot z) D_\phi(\Phi_{d;n,r}(\circ, y))(z) d\mu_d^*(z). \]

Discretization leads to a (pre-fabricated) zonal function network of the form \( \sum_k a_k(y) \phi(x \cdot z_k) \), where:

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

TINN for PINN?

---

*Mhaskar, Narcowich, Ward, 1999, Mhaskar, 2006, 2010, 2019, 2020, 2020*
Thank you.