Exploiting Lipschitz Continuity for the Kolmogorov Superposition Theorem

Jonas Actor

Rice University

SGA 2018

Contact: jonasactor@rice.edu

Dimension Reduction Methods

- ANOVA:

$$f(x_1, \dots, x_d) \approx f_0 + \sum_i f_i(x_i) + \sum_{i<j} f_{i,j}(x_i, x_j) + \dots$$

- RBF:

$$f(x_1, \dots, x_d) \approx \sum_j w_j \phi(\|\mathbf{x} - \mathbf{x}_j\|)$$

- RKHS:

$$f(x_1, \dots, x_d) = \sum_i a_i K(x, x_i)$$

Kolmogorov Superposition Theorem[1]

Any continuous $f : [0,1]^d \to \mathbb{R}$ can be written as

$$f(x_1, \ldots, x_d) = \sum_{q=0}^{2d} \chi\left(\sum_{p=1}^{d} \psi_{p,q}(x_p) \right)$$

[1] Kolmogorov, Dokl. Akad. Nauk SSSR 114:5, 1957

Implication

$$\Psi^q(x_1, \ldots, x_d) = \sum_{p=1}^{d} \psi_{p,q}(x_p)$$

(independent of f)

$$\text{KST: } f \longmapsto \chi$$

Outline of Talk

1. Constructive KST
2. Smoothness Concerns
3. Outer Function Approximation

Table of Contents

Concept

$$\Psi = (\Psi^0, \dots, \Psi^{2d}) \text{ embeds } [0,1]^d \text{ into } \mathbb{R}^{2d+1}$$

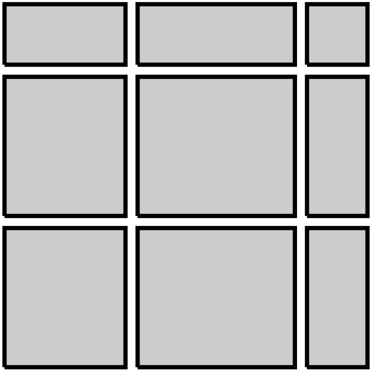
Ψ balances continuity vs. ability to discriminate points

χ assigns values of embedded space $\Psi([0,1]^d)$ to match f

Concept: Spatial Decomposition

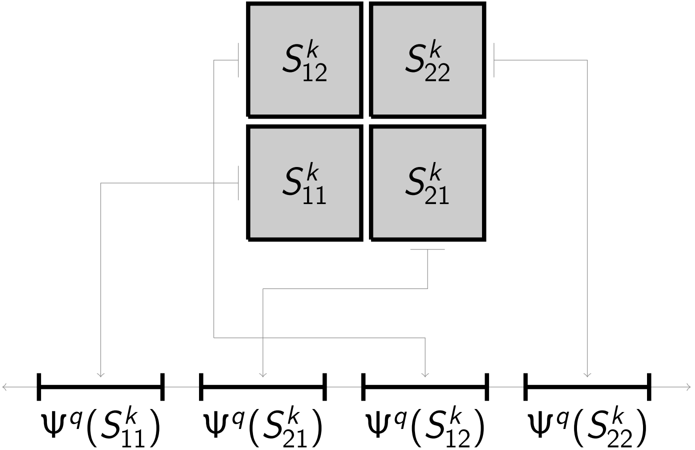
$\mathcal{S}^k$: Near partition of $[0,1]^d$

$\mathrm{diam}(\mathcal{S}^k) \to 0$ as $k \to \infty$



Disjoint Images of Squares

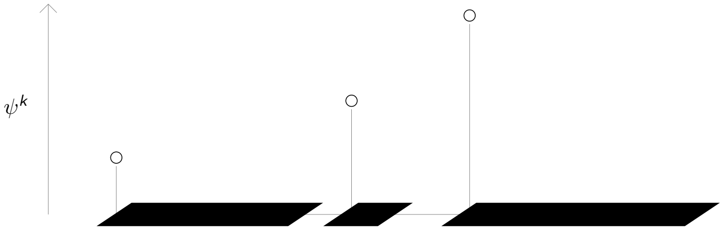
For any two squares S, $S' \in \mathcal{S}^k$, $\Psi^q(S) \cap \Psi^q(S') = \emptyset$.



Inner Function ψ

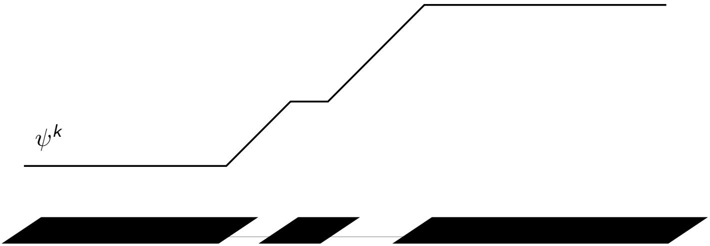
At each refinement level k:

- Assign a value of ψ^k on lower left corner of each square
- Value is fixed for all future k



Inner Function ψ

ψ^k (near) constant on squares, linear on gaps



large gaps $\rightarrow$ easier for χ to tell values of f apart

small gaps $\rightarrow$ steep ψ^k $\rightarrow$ worse oscillations

Gap Sizes

Gaps cannot be too large

Need every point $\mathbf{x} \in [0, 1]^d$ to be contained in some square for **more than half** of the sets of squares

Otherwise, not enough 'correct' information to reconstruct f

Sprecher's Reduction[2]

Define 1 function ψ instead of $2d^2 + d$ functions $\psi_{p,q}$:

$$\psi_{p,q}(x_p) = \lambda_p \, \psi(x_p + q\varepsilon)$$

$\lambda_1, \ldots, \lambda_d$ integrally independent

[2] Sprecher, Trans. AMS, 115:340–355, 1965

Cartesian product of 1D family of intervals

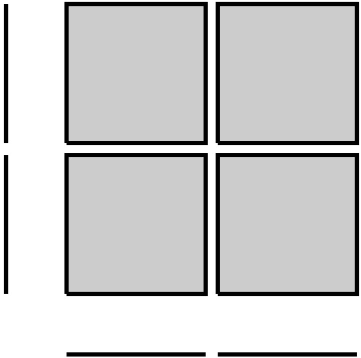


Table of Contents

Why smoothness is a concern...

Traded smoothness for variables:

- KST not feasible[3] for $\psi_{p,q} \in C^1([0,1])$
- Current $\psi_{p,q} \in \text{Hölder}([0,1])$
- Possible $\psi_{p,q} \in \text{Lip}([0,1])$

[3] Vituskin, DAN, 95:701–704, 1954.

Hölder Continous $\widehat{\psi}$

- Define iteratively: $\widehat{\psi} = \lim_{k\to} \widehat{\psi}^k$
- Fix values at points with k digits in base γ expansion[4]
  - Small increase for most points
  - Large increase for expansions ending in $\gamma - 1$
- Linearly interpolate between fixed values

[4] Sprecher, J. Constr. Approx. 1995

Hölder Continuous $\widehat{\psi}$: Setup [5] [6]

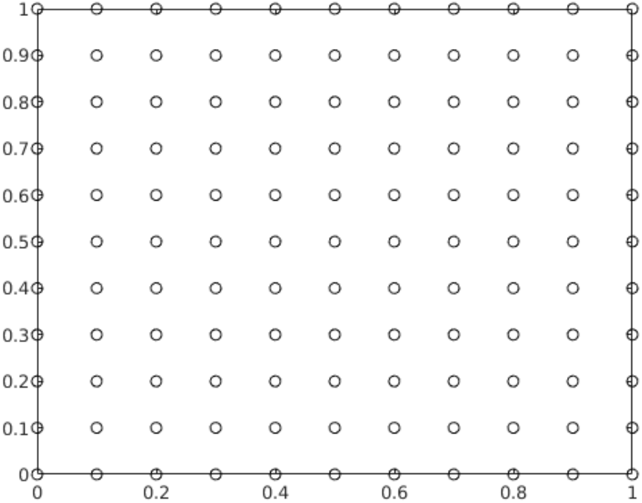
$$\text{Radix } \gamma \geq 2d + 1 \qquad\qquad \beta(k) = \tfrac{n^k - 1}{n-1}$$

$$\begin{aligned}\mathcal{D}^k &= \left\{ \frac{i}{\gamma^k} \;:\; i = 0, \dots, \gamma^k \right\} \\ &= \left\{ i_0.i_1 i_2 \dots i_k \;:\; i_\ell \in \{0, \dots, \gamma - 1\}, \ell = 0, \dots, k \right\}.\end{aligned}$$

[5] Köppen, ICANN 2002, LNCS 2415, 2002

[6] Braun and Griebel, Constr. Approx., 30:653–675, 2009

Hölder Continuous $\widehat{\psi}$: Grid



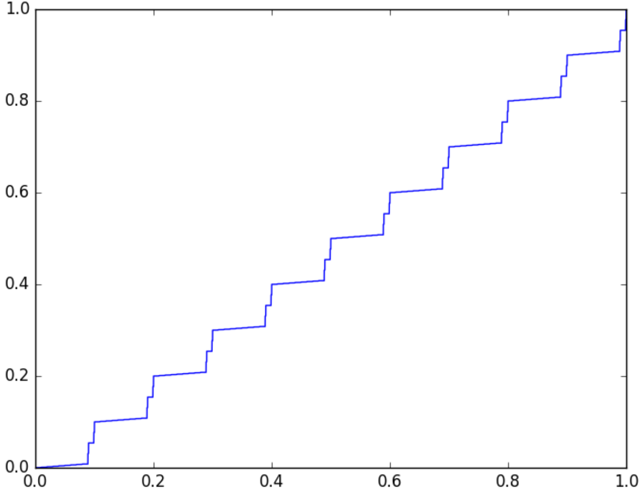
Hölder Continuous $\widehat{\psi}$: Definition

$$\widehat{\psi}^k(d_k) = \begin{cases} d_k & k = 1 \\ \widehat{\psi}^{k-1}\left(d_k - \frac{i_k}{\gamma^k}\right) + \frac{i_k}{\gamma^{\beta(k)}} & k > 1,\ i_k < \gamma - 1 \\ \frac{1}{2}\left(\widehat{\psi}^k\left(d_k - \frac{1}{\gamma^k}\right) + \widehat{\psi}^{k-1}\left(d_k + \frac{1}{\gamma^k}\right)\right) & k > 1,\ i_k = \gamma - 1 \end{cases}$$

Interpolate linearly to extend $\widehat{\psi}^k$ from $\mathcal{D}^k$ to $[0, 1]$

$$\widehat{\psi} = \lim_{k\to\infty} \widehat{\psi}^k$$

Köppen's KST Inner Function $\widehat{\psi}^7$



Need Better Than Hölder

For $\psi \in$Hölder$_\alpha([0,1])$,

$$|\psi(x) - \psi(y)| < \epsilon \quad \text{if} \quad |x - y| < \epsilon^{1/\alpha}$$

ψ is not locally Lipschitz on **any** open interval $I \subset [0,1]$

Reasoning for Lipschitz Construction

KST inner functions are strictly monotonic increasing

…which are of Bounded Variation

…so they define rectifiable curves

…which have Lipschitz reparameterizations.

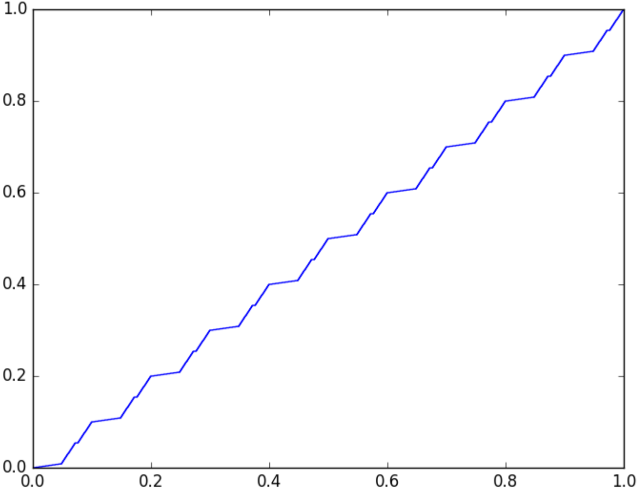
Lipschitz Reparameterization

Reparameterization $\sigma : [0, 1] \to [0, 1]$

$$\sigma(x) = \frac{\text{arclength of } \widehat{\psi} \text{ from 0 to } x}{\text{total arclength of } \widehat{\psi} \text{ from 0 to 1}}$$

$$\psi(x) = \widehat{\psi}(\sigma^{-1}(x))$$

Lipschitz Reparameterization ψ^7



Analysis

- $\psi^k = \widehat{\psi^k} \circ (\sigma^k)^{-1}$ converges uniformly to ψ
- ψ meets criteria for Kolmogorov inner function
- $\psi \in \mathrm{Lip}_2([0,1])$

Comparing $\widehat{\mathcal{S}}^k$ and $\mathcal{S}^k$

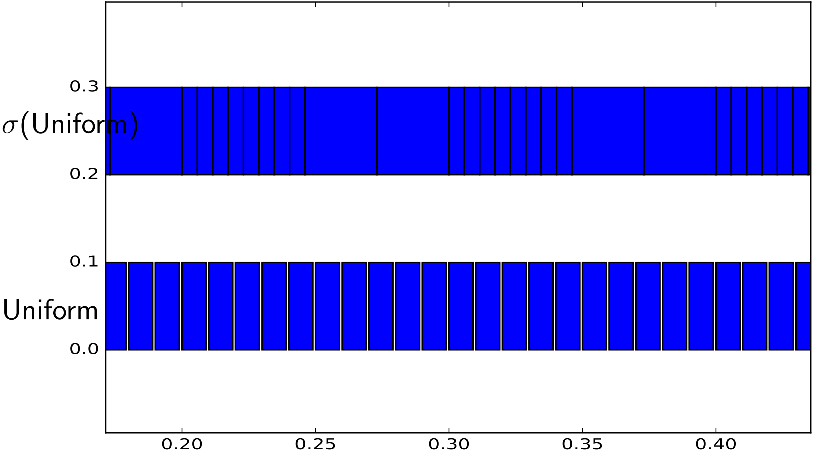


Table of Contents

Outer Function: 1D Problem

$$\chi = \lim_{r \to \infty} \chi_r \qquad f_r = f_{r-1} - \sum_{q=0}^{2d} \chi_{r-1} \circ \Psi^q$$

For $r \in \mathbb{N}$, choose k_r so that the oscillation of

$$f_{r-1} - \sum_{q=0}^{2d} \chi_{r-1} \circ \Psi^q$$

is bounded by $\frac{\|f_{r-1}\|_\infty}{d+1}$

Outer Function: 1D Problem

$$\chi_r : \mathbb{R} \to \mathbb{R} \text{ interpolates}$$

$$\left\{ \left(\Psi^q(\mathbf{d_{k_r}}), \frac{1}{d+1} f(\mathbf{d_{k_r}}) \right) \ : \ \mathbf{d_{k_r}} \in \prod_{p=1}^{d} \sigma(\mathcal{D}_{k_r} + q\varepsilon) \right\}$$

Adaptive Approximation

Number of interpolation points grows exponentially with d at each level k_r

Functions with high oscillation on squares require quick growth of k_r wrt. r

Need to be smart about choosing interpolation points

Adaptive Approximation

For Hölder ψ, only need polynomial $P(d)$ points [7]

If oscillation too large at step k_r, choose k_{r+1} so that oscillation is controlled locally

Stabilized by Lipschitz continuity of Ψ

[7] Griebel and Braun, ICHPSC 2009

Conclusions

- KST leverages superpositions for potential dimension reduction
- Lipschitz inner function provides for practical KST computation
- Look to reduce number of interpolation points for outer function

Next Steps

Free Interpolation condition on χ

Adaptive Outer function theory

Framework for computing outer function χ

- Requires squares at each level k
- Requires final high-resolution Ψ
- Requires accurate measure of oscillation