# Understanding Multivariate Computation using the Kolmogorov Superposition Theorem

Matthew Knepley and Jonas Actor

Computer Science and Engineering
University at Buffalo

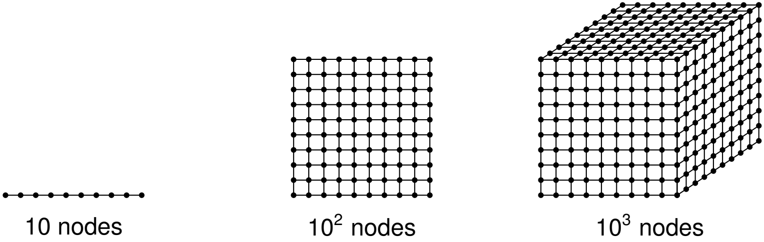
MCS Seminar, Argonne National Laboratory
Chicago, IL     September 19, 2018

# Collaborators

KST Representation Collaboration

Jonas Actor

# Curse of Dimensionality




Cost of simulation / computation grows **exponentially**

# Main Question

Do functions of three variables exist at all?[1]

[1] Pólya and Szegö, Problems and Theorems of Analysis, 1925 (German), transl. 1945, reprinted 1978

**Q**: Can any function of three variables be expressed using functions of only two variables?[2]

[2] Hilbert, Göttinger Nachrichten, 1900

[3] Arnol'd, Dokl. Akad. Nauk SSSR 114:5, 1957

# Nomography

**Q**: Can any function of three variables be expressed using functions of only two variables?[2]

Ex: Cardano's Formula for roots of a cubic equation

[2] Hilbert, Göttinger Nachrichten, 1900

[3] Arnol'd, Dokl. Akad. Nauk SSSR 114:5, 1957

# Nomography

**Q**: Can any function of three variables be expressed using functions of only two variables?[2]

Ex: Cardano's Formula for roots of a cubic equation

**A**: Any *continuous* function of three variables can be expressed using *continuous* functions of only two variables.[3]

[2] Hilbert, Göttinger Nachrichten, 1900

[3] Arnol'd, Dokl. Akad. Nauk SSSR 114:5, 1957

**Q**: Can any function of three variables be expressed using only univariate functions and addition?

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

# Nomography

**Q**: Can any function of three variables be expressed using only univariate functions and addition?

**A**: Yes[4], for any continuous $f : [0,1]^n \to \mathbb{R}$

$$f(x_1, \dots, x_n) = \sum_{q=0}^{2n} \chi \left( \sum_{p=1}^{n} \psi_{p,q}(x_p) \right).$$

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

# What's going on here?

$$f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} \chi_q \left( \sum_{p=1}^{n} \psi_{p,q}(x_p) \right)$$

# What's going on here?

$$f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} \chi_q \left( \sum_{p=1}^{n} \psi_{p,q}(x_p) \right)$$

Original function $f : [0, 1]^n \to \mathbb{R}$

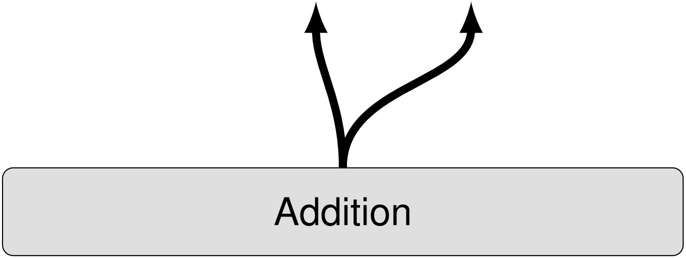
# What's going on here?

$$f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} \chi_q \left( \sum_{p=1}^{n} \underbrace{\psi_{p,q}(x_p)} \right)$$

Inner function $\psi_{p,q} : [0,1] \to \mathbb{R}$

# What's going on here?

$$f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} \chi_q \left( \sum_{p=1}^{n} \psi_{p,q}(x_p) \right)$$



# What's going on here?

$$f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} \chi_q \left( \sum_{p=1}^{n} \psi_{p,q}(x_p) \right)$$

Outer function $\chi : \mathbb{R} \to \mathbb{R}$

# What's going on here?

$$f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} \chi_q \left( \sum_{p=1}^{n} \psi_{p,q}(x_p) \right)$$

Function composition

# 2D: $n = 2$

$$
\begin{aligned}
f(x, y) = \chi_0\big(\psi_{x,0}(x) + \psi_{y,0}(y)\big) \\
+ \chi_1\big(\psi_{x,1}(x) + \psi_{y,1}(y)\big) \\
+ \chi_2\big(\psi_{x,2}(x) + \psi_{y,2}(y)\big) \\
+ \chi_3\big(\psi_{x,3}(x) + \psi_{y,3}(y)\big) \\
+ \chi_4\big(\psi_{x,4}(x) + \psi_{y,4}(y)\big)
\end{aligned}
$$

# 2D: $n = 2$

$$
\begin{aligned}
f(x,y) = \chi_0\big(\lambda_x\psi(x) &+ \lambda_y\psi(y)\big) \\
+ \chi_1\big(\lambda_x\psi(x+\epsilon) &+ \lambda_y\psi(y+\epsilon)\big) \\
+ \chi_2\big(\lambda_x\psi(x+2\epsilon) &+ \lambda_y\psi(y+2\epsilon)\big) \\
+ \chi_3\big(\lambda_x\psi(x+3\epsilon) &+ \lambda_y\psi(y+3\epsilon)\big) \\
+ \chi_4\big(\lambda_x\psi(x+4\epsilon) &+ \lambda_y\psi(y+4\epsilon)\big)
\end{aligned}
$$

Single $\psi$ function (Sprecher)

# 2D: $n = 2$

$$
\begin{aligned}
f(x,y) = {} & \chi\big(\lambda_x\psi(x) + \lambda_y\psi(y)\big) \\
& + \chi\big(\lambda_x\psi(x+\epsilon) + \lambda_y\psi(y+\epsilon) + \delta\big) \\
& + \chi\big(\lambda_x\psi(x+2\epsilon) + \lambda_y\psi(y+2\epsilon) + 2\delta\big) \\
& + \chi\big(\lambda_x\psi(x+3\epsilon) + \lambda_y\psi(y+3\epsilon) + 3\delta\big) \\
& + \chi\big(\lambda_x\psi(x+4\epsilon) + \lambda_y\psi(y+4\epsilon) + 4\delta\big)
\end{aligned}
$$

Single $\chi$ function (Lorentz) and $\psi$ function (Sprecher)

# Universality

$$\Psi^q(x_1, \dots, x_n) = \sum_{p=1}^{n} \psi_{p,q}(x_p)$$

is independent of $f$, so that

$$KST : f \to \chi$$

# Outline

# Topographic KST

$$f(x, y) = \text{ elevation in Chugash Mountains at } (x, y)$$


https://viewer.nationalmap.gov/basic/

# Topographic KST

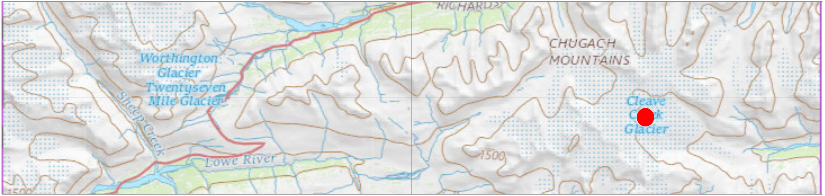
$$f(\text{Cleave Creek Glacier}) = \; ?$$


https://viewer.nationalmap.gov/basic/

# Topographic KST

A B

1
2



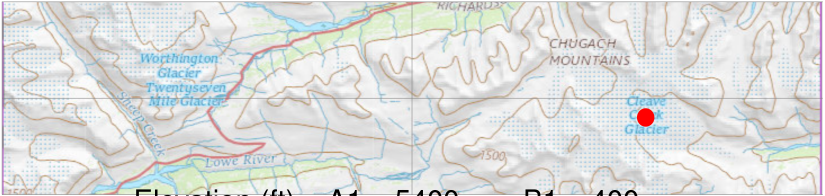

| | | | |
|---|---|---|---|
| Tsina Glacier | A1 | Tsina River | B1 |
| 27 Mile Glacier | A2 | Chugash Mtns. | B2 |

# Topographic KST

A B

1

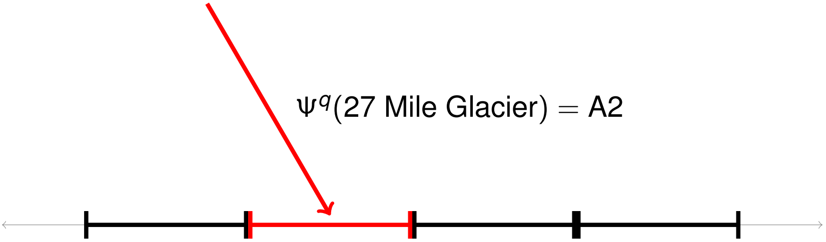
2




| | | | | |
|---|---|---|---|---|
| Elevation (ft) | A1 | 5400 | B1 | 400 |
| | A2 | 4800 | B2 | 5700 |

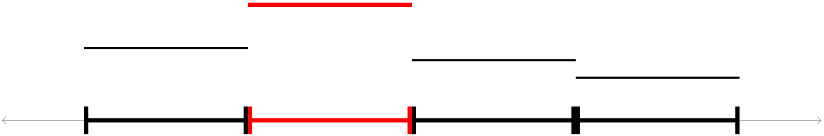
# Topographic KST


$\Psi^q(\text{27 Mile Glacier}) = \text{A2}$



# Topographic KST


$$f(\text{27 Mile Glacier}) = \chi(\text{A2})$$



# Topographic KST

A B

1

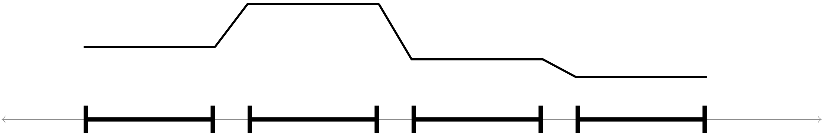


2


# Topographic KST

A B

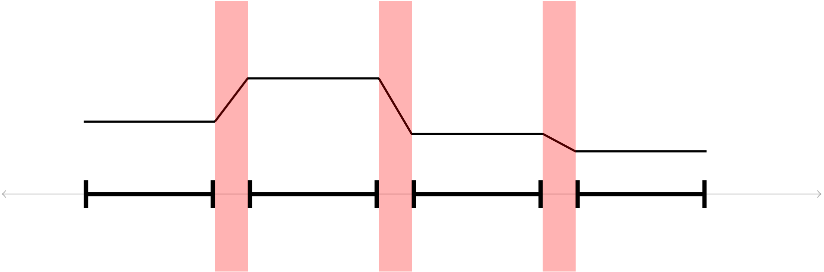
1 2

# Topographic KST

A B

1
2

# Outline

# Kolmogorov Strategy

- Approximate with *constants*
- Insert *gaps* to preserve continuity
- *Duplicate and shift* to cover domain

# Kolmogorov Strategy

- Approximate with *constants*
- Insert *gaps* to preserve continuity
- *Duplicate and shift* to cover domain

$\Psi$: balances continuity against discriminating between different points

# Kolmogorov Strategy

- Approximate with *constants*
- Insert *gaps* to preserve continuity
- *Duplicate and shift* to cover domain

$\Psi$: regularity vs. point separation

$\chi$ assigns values to shifted sums of our $R^{2n+1}$ embedding to match $f$

# Kolmogorov Strategy

- Approximate with *constants*
- Insert *gaps* to preserve continuity
- *Duplicate and shift* to cover domain

$\Psi$: regularity vs. point separation
$\chi$: approximates $f$

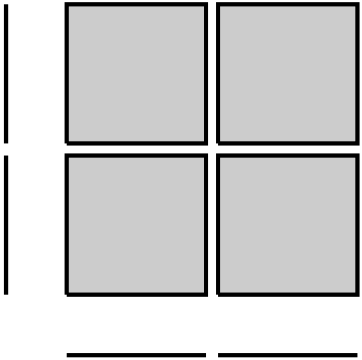
# Outline

2 Abstract KST
- Spatial Decomposition
- Inner Functions

# Spatial Decomposition

## Single town

Cartesian product $\mathcal{S}$ of a set of intervals $\mathcal{I}$ that nearly cover $[0, 1]$

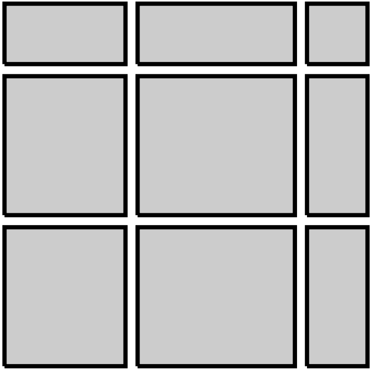


# Spatial Decomposition

## Refined towns

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

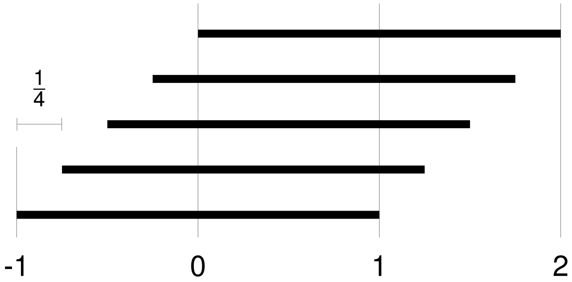
$$\text{diam}(\mathcal{S}^k) \to 0 \text{ as } k \to \infty$$



# Spatial Decomposition

## Cover gaps

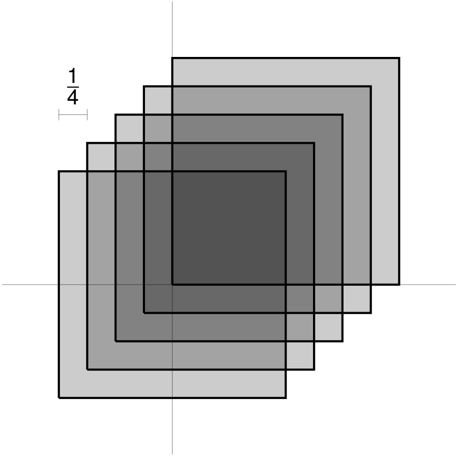
$2n+1$ copies $\mathcal{I}_q$ of the same set of intervals $\mathcal{I}$ shifted by $\epsilon$



# Spatial Decomposition

## Overlapping towns

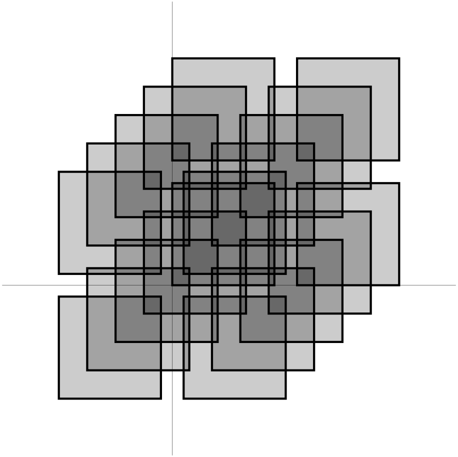
Cartesian product $\mathcal{S}_q$ of shifted intervals $\mathcal{I}_q$



# Spatial Decomposition

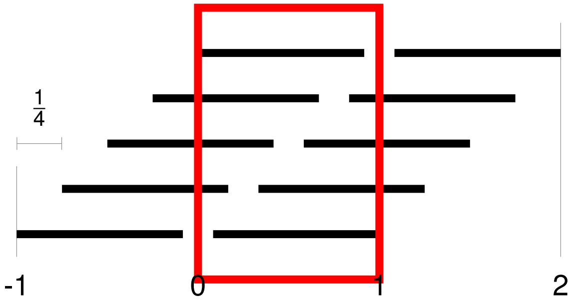
## Refined overlapping towns

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



# Gap Requirement

## Every $x \in [0, 1]$ in **All But One** of $\mathcal{I}_q$



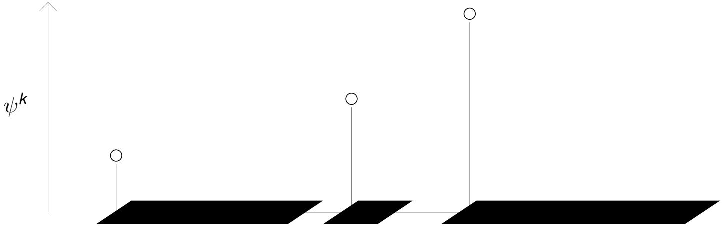

**All But One** of the intervals implies
**More than Half** of the squares

# Outline

# Inner Function $\psi$

At each refinement level $k$:

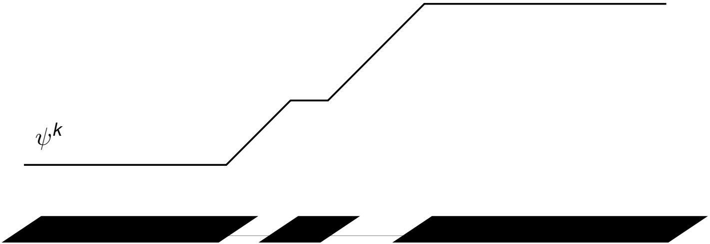
- Assign a value of $\psi^k$ at the lower left corner of each square
- Value is fixed for all future $k$



# Inner Function $\psi$

$\psi^k$ (near) constant on squares, linear on gaps

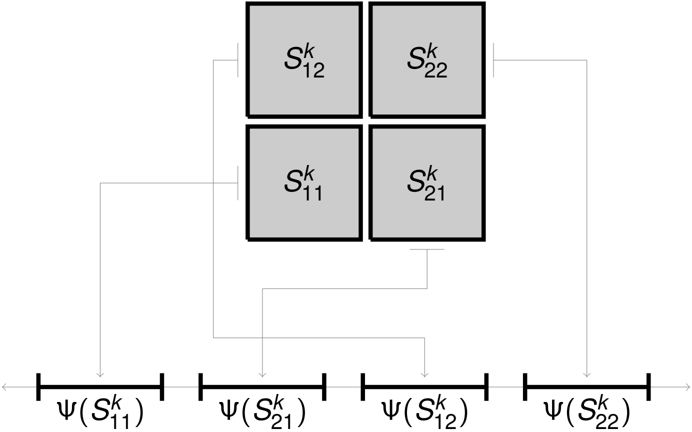
$\psi = \lim_{k\to\infty} \psi^k$ uniformly




Small gaps $\to$ steep $\psi^k$, but
Large gaps can violate **All But One**

# Disjoint Image Condition

$$\forall S,\ S' \in \mathcal{S}^k, \Psi(S) \cap \Psi(S') = \emptyset$$



# Kolmogorov Requirements for the Inner Function

**Refinement:**

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

**More Than Half:**

$\forall x \in [0,1]^n, x \in \mathcal{S}_q^k$ for at least $n+1$ of the $q$'s

**Disjoint Image:**

$\forall S,\ S' \in \mathcal{S}_q^k, \Psi(S) \cap \Psi(S') = \emptyset$

**Monotonicity:**

Function $\psi$ is strictly monotonic increasing

# Regularity of Inner Function

KST trades smoothness for variables

- KST not feasible for $\psi_{p,q} \in C^1([0,1])$[5]
- Possible to construct $\psi_{p,q} \in \mathrm{Lip}([0,1])$[6]
- Only known construction $\psi_{p,q} \in \mathrm{Hölder}([0,1])$

[5] Vituskin, DAN, 95:701–704, 1954.
[6] Fridman, DAN, 177:1019–1022, 1967.

# Outline

# Outline

3 Concrete KST
- Sprecher Construction
- Lipschitz Construction
- Outer Functions

# Sprecher Construction $\widehat{\psi}$ [7]

- Inner function $\widehat{\psi}$ is the uniform limit $\lim_{k \to \infty} \widehat{\psi}^k$
- Fix values at numbers with $k$ digits in base-$\gamma$ expansion
  - Almost flat for most points
  - Large increase for expansions ending in $\gamma - 1$
- Linearly interpolate between fixed values

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

# Sprecher Construction $\widehat{\psi}$ [8]

$$\text{Radix } \gamma \geq 2d + 1$$

$$\mathcal{D}^k = \left\{ \frac{i}{\gamma^k} \quad : \quad i = 0, \ldots, \gamma^k \right\}$$
$$= \left\{ 0.i_0 i_1 \ldots i_k \quad : \quad i_\ell \in [0, \gamma - 1], \ \ell \in [0, k] \right\}$$

$$\beta(k) = \frac{n^k - 1}{n - 1}$$

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

# Köppen Construction $\widehat{\psi}$ [9]

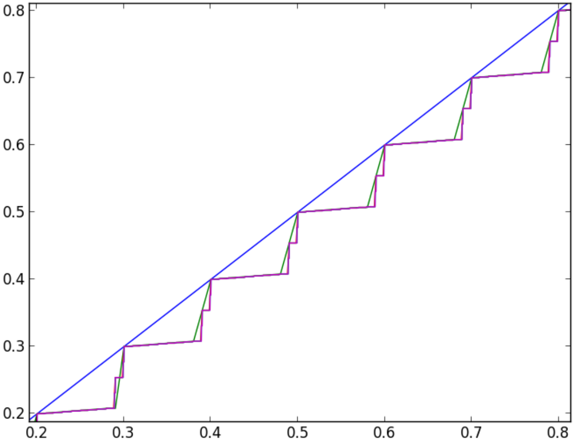
$$\widehat{\psi}^k(d_k) = \begin{cases} d_k & k = 1 \\ \widehat{\psi}^{k-1}\left(d_k - \frac{i_k}{\gamma^k}\right) + \frac{i_k}{\gamma^{\beta(k)}} & k > 1,\ i_k < \gamma - 1 \\ \frac{1}{2}\left(\widehat{\psi}^k\left(d_k - \frac{1}{\gamma^k}\right) + \widehat{\psi}^{k-1}\left(d_k + \frac{1}{\gamma^k}\right)\right)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

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

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

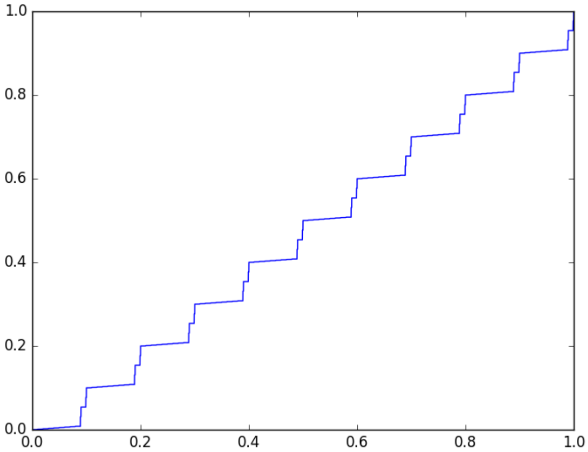
[9] Köppen, ICANN 2002, LNCS 2415, 2002

# Köppen Construction $\widehat{\psi}$ [10]




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

# Köppen Construction $\widehat{\psi}$[11], $k = 7$




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

# Köppen Construction $\widehat{\psi}$

## Approximability

While $\widehat{\psi}$ satisfies all Kolomogorov requirements[12], it's not locally Lipschitz on **any** open interval $I \subset [0, 1]$.

In fact, $\psi \in \text{Hölder}_{\alpha}([0, 1])$ for $\alpha = \log_{2n+2} 2$,

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

For $n = 2$, $\alpha^{-1} \approx 2.5$, but

for $n = 10$ we have $\alpha^{-1} \approx 4.5$.

[12] Braun and Griebel, Constr. Approx., 30:653-675, 2009

# Outline

3 Concrete KST
- Sprecher Construction
- Lipschitz Construction
- Outer Functions

# Reasoning for Lipschitz Construction

KST inner functions are

. . . strictly monotonically increasing

. . . so they are of Bounded Variation

. . . so they define rectifiable curves

. . . which have Lipschitz reparameterizations.

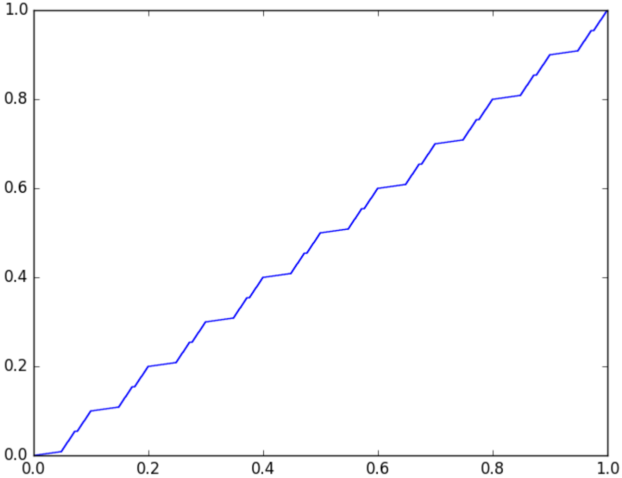
# Lipschitz Reparameterization

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

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

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

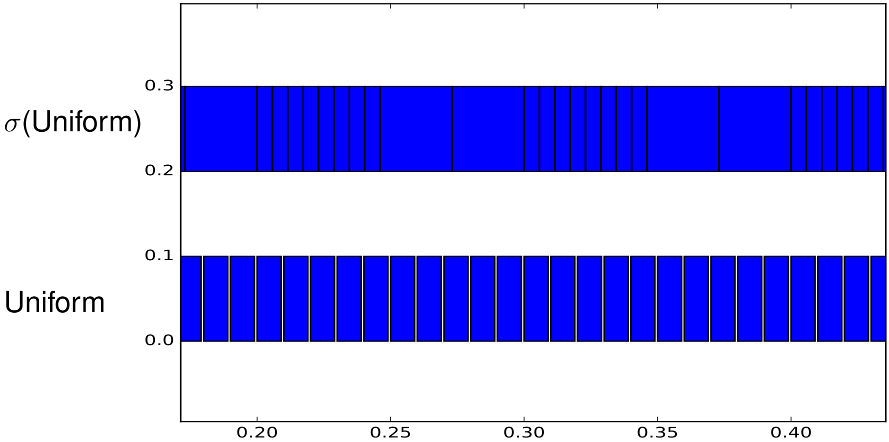
# Lipschitz Reparameterization $\psi$, $k = 7$



# KST Proof with Lipschitz $\psi$

- $\psi^k = \widehat{\psi^k} \circ (\sigma^k)^{-1}$ converges uniformly to $\psi$
- $\psi$ satisfies the Kolmogorov requirements
- $\psi \in \mathrm{Lip}_2([0,1])$

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



# Outline

# Kolmogorov Outer Function

Define a residual

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

and choose $k_r$, with $r \in \mathcal{N}$, so that the oscillation of

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

is bounded by $\frac{\|f_{r-1}\|_\infty}{d+1}$. Then by a Fixed Point Theorem,

$$\chi = \lim_{r \to \infty} \chi_r.$$

# Outer Function Construction

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

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

# Outline

# Conclusions

- KST leverages superpositions for dimension reduction
- Practical KST computation requires Lipschitz inner functions
- Need to develop adaptive theory for outer function construction

# Thank You!

http://cse.buffalo.edu/~knepley