Lipschitz Inner Functions in Kolmogorov’s Superposition Theorem

Jonas Actor

Rice University

13 September 2017
# Table of Contents

1. Motivation
2. Characteristics of Inner Functions
3. Construction of a Lipschitz Inner Function
4. Implementation and Results
5. Moving Forward
1 Motivation

2 Characteristics of Inner Functions

3 Construction of a Lipschitz Inner Function

4 Implementation and Results

5 Moving Forward
Hilbert’s 13th Problem

Problem Statement (1900)

Can the solution $x$ to the 7th degree polynomial equation

$$x^7 + ax^3 + bx^2 + cx + 1 = 0$$

be represented by a finite number of compositions of bivariate continuous/analytic functions using the three variables $a, b, c$?
Larger focus to Hilbert’s Problem:

- Which functions can be represented using a finite number of compositions of simpler functions?
  - Continuous functions $\rightarrow$ Continuous functions?
  - Analytic functions $\rightarrow$ Continuous functions?
  - Analytic functions $\rightarrow$ Analytic functions?

- How ‘simple’ or ‘complex’ is a function?
  - Number of variables?
  - Some other measure
An Example

\[ f(x_1, x_2) = x_1x_2 \]

Can we create \( \phi_1, \phi_2, \psi_{11}, \psi_{12}, \psi_{21}, \psi_{22} \) univariate polynomials such that

\[ f(x_1, x_2) = \sum_{i=1}^{2} \phi_i \circ \sum_{j=1}^{2} \psi_{ij}(x_j)? \]
An Example

Let

\[ \phi_1(z) = \frac{1}{4}z^2 \quad \phi_2(z) = -\frac{1}{4}z^2, \]

and

\[ \psi_{11}(x_1) = x_1 \quad \psi_{12}(x_2) = x_2 \]
\[ \psi_{21}(x_1) = x_1 \quad \psi_{22}(x_2) = -x_2 \]

Then,

\[
\sum_{i=1}^{2} \phi_i \left( \sum_{j=1}^{2} \psi_{ij}(x_j) \right) = \frac{1}{4}(x_1 + x_2)^2 - \frac{1}{4}(x_1 - x_2)^2
\]

\[
= \frac{1}{4}(x_1^2 + 2x_1x_2 + x_2^2) - \frac{1}{4}(x_1^2 - 2x_1x_2 + x_2^2)
\]

\[
= \frac{1}{4}(2x_1x_2 + 2x_1x_2)
\]

\[
= x_1x_2.
\]
Can we do better?
  - Use fewer terms!

What concessions do we make?
  - Give up using polynomials, instead continuous functions
An Example

Let

\[ \phi_1(z) = \exp(z) \quad \psi_i(x_i) = \log(x_i) \]

Then,

\[
\phi_1 \left( \sum_{j=1}^{2} \psi_j(x_j) \right) = \exp(\log(x_1) + \log(x_2)) \\
= \exp(\log(x_1 x_2)) \\
= x_1 x_2.
\]

We have reduced the number of outer summands by 1.
An Example

- Representation not unique
- Number of terms in (inner) summand depends on dimension
- Traded ‘complexity’ of functions used
- Note, these are all very special functions
Hilbert conjectured the answer was ‘no’, that even for continuous functions, such a representation was not always possible.

A. Kolmogorov and V.I. Arnold made progress in the 1950s.

Arnold (at age 19) proved the answer was ‘yes’ in 1957: any multivariate continuous function can be represented as a superposition of bivariate continuous functions.

Two weeks later, Kolmogorov reduced the bivariate functions from Arnold to univariate functions.
Kolmogorov’s Superposition Theorem (KST) (1957)

Theorem

Let \( f : \mathbb{R}^n \rightarrow \mathbb{R} \in C([0,1]^n) \) where \( n \geq 2 \). Then, there exist \( \psi^{pq} : [0,1] \rightarrow \mathbb{R} \in C[0,1] \) and \( \chi_q : \mathbb{R} \rightarrow \mathbb{R} \in C(\mathbb{R}) \), where \( p \in \{1, \ldots, n\} \) and \( q \in \{0, \ldots, 2n\} \), such that

\[
f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} \chi^q \left( \sum_{p=1}^{n} \psi^{pq}(x_p) \right).
\]
Kolmogorov’s Superposition Theorem: Reformulation

This reformulation is due to Sprecher.

**Theorem**

Let \( f : \mathbb{R}^n \to \mathbb{R} \in C([0, 1]^n) \) where \( n \geq 2 \). Fix \( \epsilon \leq \frac{1}{2^n} \), and choose \( \lambda \in \mathbb{R} \) such that \( 1 = \lambda^0, \lambda^1, \ldots, \lambda^{n-1} \) are integrally independent. Then, there exist \( \psi : [-1, 1] \to \mathbb{R} \in C[-1, 1] \) and \( \chi_q : \mathbb{R} \to \mathbb{R} \in C(\mathbb{R}) \) for \( q \in \{0, \ldots, 2n\} \), such that

\[
f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} \chi^q \left( \sum_{p=1}^{n} \lambda^p \psi(x_p + q\epsilon) \right).
\]
Implications

Multivariate functions suffer from the ‘curse of dimensionality’, making computation hard for higher dimensions.

Multivariate continuous functions are really

\[
\begin{align*}
\text{univariate continuous functions} \\
\text{addition} \\
\text{function composition}
\end{align*}
\]

We understand each of those three things very well...
Quest: Can we use KST to represent multivariate functions for efficient computation?
Table of Contents

1 Motivation

2 Characteristics of Inner Functions

3 Construction of a Lipschitz Inner Function

4 Implementation and Results

5 Moving Forward
Constructions are by induction on $j \in \mathbb{N}$

Throughout the rest of this talk, fix $\epsilon = 1/2n$
Town: closed interval
System of towns: set of disjoint closed intervals

- $T_j$ a system of towns $\subseteq [-1, 1]$
- $T_j^q$ a system of towns $\subseteq [-1 + q\epsilon, 1 + q\epsilon]$ where

$$T_j^q = \{ t + q\epsilon : t \in T_j \}.$$

- Enumerate the towns in $T_j^q$ by indices $1 \leq i \leq m_j$

$$T_j^q = \{ t_1^q, t_2^q, \ldots, t_m^q \}.$$
Space Partitioning

Squares: products of towns

- \( S^q_{j;i_1,\ldots,i_n} = \prod_{p=1}^{n} t^q_{i_p} \) for towns \( t^q_{i_p} \in T^q_j \) for \( p = 1, \ldots, n \)
- \( \mathcal{I}^q_j = \{ S^q_{j;i_1,\ldots,i_n} : 1 \leq i_1, \ldots, i_n \leq m_j \} \) set of all squares

Squares are pairwise disjoint: for any \( q \), if \( (i_1, \ldots, i_n) \neq (i'_1, \ldots, i'_n) \), then
\[
S^q_{j;i_1,\ldots,i_n} \cap S^q_{j;i'_1,\ldots,i'_n} = \emptyset.
\]
Define $\Psi^q(x_1, \ldots, x_n) = \sum_{p=1}^n \psi^{pq}(x_p)$ for each $q \in \{0, \ldots, 2n\}$, where $\psi^{pq} \in C[0, 1]$

**Lemma**

For each $j, q$, suppose the families of squares $S_j^q$ satisfy the following:

1. Each point $x \in [0, 1]^n$ intersects squares from at least $n + 1$ of the $2n + 1$ families of squares
2. $\sup_{S_j^q, i_1, \ldots, i_n \in S_j^q} \text{Diam}[S_j^q] \rightarrow 0$ uniformly as $j \rightarrow \infty$
3. $\Psi^q(S_j^q, i_1, \ldots, i_n) \cap \Psi^q(S_j^q, i'_1, \ldots, i'_n) = \emptyset$ when $(i_1, \ldots, i_n) \neq (i'_1, \ldots, i'_n)$.

Then, any function $f \in C([0, 1]^n)$ admits a KST representation

$$f = \sum_{q=0}^{2n} \chi^q \circ \Psi^q.$$
\begin{itemize}
  \item $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ integrally independent
  \item $\psi_j : [-1, 1] \rightarrow \mathbb{R}$
  \item $\psi := \lim_{j \rightarrow \infty} \psi_j$ uniformly
  \item For $q \in \{0, \ldots, 2n\}$ define $\Psi^q : [0, 1]^n \rightarrow \mathbb{R}$
    \[ \Psi^q(x_1, \ldots, x_n) = \sum_{p=1}^{n} \lambda_p \psi(x_p + q\epsilon). \]
\end{itemize}
Lemma: Reformulation

For each $j, q$, suppose the systems of towns $\mathcal{T}_j^q$ and function $\psi_j$ satisfy the following:

1. Each point $x \in [0, 1]$ intersects towns from at least $2n$ of the $2n + 1$ systems of towns
2. $\sup_{t \in \mathcal{T}_j} \text{Diam}(t) \to 0$ uniformly as $j \to \infty$
3. $\psi_j(t_1) \cap \psi_j(t_2) = \emptyset$ but are rational for any $t_1, t_2 \in \mathcal{T}_{j'}$ for $j' \leq j$

Then, any function $f \in C([0, 1]^n)$ admits a KST representation

$$f = \sum_{q=0}^{2n} \chi^q \circ \psi^q.$$
Construction of squares (specifically, the $n + 1$ of $2n + 1$ property and the shrinking diameter) are important for outer function.

Hard part of inner function construction is the pairwise disjoint image condition.
Smoothness of Inner Functions

- Original KST Proof (1957): Hölder continuous
  - Squares are uniformly spaced, and scale self-similarly
- Fridman (1967): Can be Lipschitz continuous, constant 1
- Vitushkin and Henkin (1954): Not differentiable at a dense set of points
Lipschitz vs. Hölder Continuity

**Definition**

A function \( f : [0, 1]^n \rightarrow \mathbb{R} \) is **Lipschitz continuous with constant** \( C \) if \( \forall x, y \in [0, 1]^n \),

\[
\| f(x) - f(y) \| \leq C \| x - y \| .
\]

**Definition**

A function \( f : [0, 1]^n \rightarrow \mathbb{R} \) is **Hölder(\( \alpha \) continuous with constant** \( C \)** for \( \alpha \in (0, 1) \), if \( \forall x, y \in [0, 1]^n \),

\[
\| f(x) - f(y) \| \leq C \| x - y \|^{\alpha} .
\]

Hölder functions suffer from high storage/evaluation complexity, making them impractical for computation
Table of Contents

1 Motivation

2 Characteristics of Inner Functions

3 Construction of a Lipschitz Inner Function

4 Implementation and Results

5 Moving Forward
Terminology

- $\mathcal{T}_j$: system of towns (closed intervals) at refinement level $j$
- $\mathcal{T}_j^q$: system of shifted towns

\[
\mathcal{T}_j^q = \left\{ t + q\varepsilon : t \in \mathcal{T}_j, q \in \{-2n, \ldots, 2n\} \right\}.
\]

- $\psi_j : [-1, 1] \rightarrow \mathbb{R}$
- $\psi = \lim_{j \rightarrow \infty} \psi_j$
- $\Xi_j : [0, 1] \rightarrow \mathbb{N}$

\[
\Xi_j(x) = \left| \left\{ q \in \{0, \ldots, 2n\} : \exists t \in \mathcal{T}_j^q \text{ such that } x \in t \right\} \right|
\]
Lemma

It is sufficient to complete KST representation, if for each $j \in \mathbb{N}$, the system of towns $\mathcal{T}_j$ and the function $\psi_j : [-1, 1] \to \mathbb{R} \in C[-1, 1]$ satisfy the following:

1. $\sup_{t \in \mathcal{T}_j} \mathrm{Diam}(t) \to 0$ uniformly as $j \to \infty$.
2. $\forall x \in [0, 1], \exists j(x) \in \{2n, 2n + 1\}$
3. $\forall t \in \mathcal{T}_j, \psi_j(t) \in \mathbb{Q}$
4. $\forall t_1 \neq t_2 \in \mathcal{T}_j, \psi(t_1) \cap \psi(t_2) = \emptyset$.
5. $\psi_j$ is piecewise linear with maximum slope of $\hat{m}_j = 1 - 2^{-j}$. 
Start with $\psi_0 \equiv 0$ and $\mathcal{T}_0 = \{[-1, 1]\}$. Then for $j \in \mathbb{N}$ do:

Select $\widehat{T}_j \subseteq T_j$ towns to break (length $\geq 2^{-j}$)
- Find Holes
- Solve for Plugs
- Create Gaps and Update
Break $t \in \tilde{T}_j$ at its midpoint $p$ by removing an open interval $g$.

$$t \mapsto t_- \cup g \cup t_+ \quad p \in g$$

**Danger**: $p$ might no longer be contained by enough systems of towns! Might cause $\Xi_{j+1}(p) = 2n - 1$. 
Figure: Sketch of scenario where a break point $p$ falls into a hole, $n = 2$. 
Solution: Add in small ‘plugs’ so that when we remove a gap around $p$, we do not lose containment.
Solve for Plugs

\[ \hat{p} \]

\[ \psi_j \]

\[ \psi_{j+1} \text{ from blue } \pi \]

\[ \psi_{j+1} \text{ from red } \pi \]

\[ \pi \]

Figure: Sketch of how size of plugs changes the slope of \( \psi_{j+1} \).

Danger: Adding in a plug \( \implies \) slope of \( \psi_{j+1} \) might exceed our bound!
Solve for Plugs

How big can $\pi$ be for a maximum slope of $\hat{m} = 1 - 2^{-j-1}$?

- Might need more than one plug per break point (no more than 2)
- Might need more than one plug per hole

**Solution:** Solve a linear system!
Solve for Plugs

- For $\nu$ break points $p_i$, $1 \leq i \leq \nu$, that shift into hole $h = (b_0, a_{\nu+1})$
- Want disjoint plugs $\pi = [a_i, b_i]$ $a_i, b_i$ unknown

\[
\hat{p}_i := p_i - q_i \epsilon \in \pi_i \subset h
\]

- $\psi_{j+1}$ monotonic, piecewise linear, constant on towns/plugs
- $\psi_j(\hat{p}_i) = \psi_{j+1}([a_i, b_i])$

We use the following notation for (known) function values:

\[
f_0 = \psi_j(b_0) \\
f_i = \psi_j(\hat{p}_i) \quad 1 \leq i \leq \nu \\
f_{\nu+1} = \psi_j(a_{\nu+1})
\]
Solve for Plugs

Figure: Sketch of scenario for finding two plugs, with $\psi_j$ in black and $\psi_{j+1}$ in blue. Note the symmetry constraint $b_1 - \hat{p}_1 = \hat{p}_2 - a_2$ is enforced.
Solve for Plugs

\[ \nu + 1 \text{ equations:} \]
\[ \hat{m}(a_i - b_{i-1}) = f_i - f_{i-1}, \quad 1 \leq i \leq \nu + 1. \]

\[ \nu - 1 \text{ symmetry constraints:} \]
\[ b_i - \hat{p}_i = \hat{p}_{i+1} - a_{i+1}, \quad 1 \leq i \leq \nu - 1. \]
Solve for Plugs

This provides the linear system $Cx = z$

- $C$ is block diagonal, invertible $\implies$ unique solution exists
- $\psi_j$ monotonic increasing $\implies$ plugs are disjoint with non-empty interior

For each $h$, update $\mathcal{T}_j$ to include the plugs $\pi_i$. 
Create Gaps

Recall our goal to ‘break’ \( t \in \widetilde{T}_j \) at break point \( p \):

\[
t \mapsto t_- \cup g \cup t_+ \quad p \in g
\]

At this point, \( \Xi_j(p) = 2n + 1 \implies \Xi_{j+1}(p) \geq 2n : \)

\[
\forall q \in \{0, \ldots, 2n\}, \exists t_q \in \widetilde{T}_j^q \text{ such that } p - q \epsilon \in t_q,
\]

so we can remove some open \( g \) from \( t \) while keeping that \( \forall x \in [0, 1], \Xi_j(x) \in \{2n, 2n + 1\} \).
Create Gaps

Figure: Sketch of creating a gap to our previous scenario.
Assign $\psi_{j+1}(t_-) = \psi_j(t)$, and choose value for $\psi_{j+1}(t_+)$ so that:

- Maintain monotonicity
- Slope is bounded $\leq \frac{1}{2}$

Update $\mathcal{T}_j$ to $\mathcal{T}_{j+1}$ by replacing $t$ with $t_-, t_+$. 

Table of Contents

1 Motivation

2 Characteristics of Inner Functions

3 Construction of a Lipschitz Inner Function

4 Implementation and Results

5 Moving Forward
Implemented in Python, in serial (for now)
Stores one system of towns $T_j$ as an Interval Tree
Extended precision
Figure: System of towns, after refinement $j = 1$
Figure: System of towns, after refinement $j = 2$
**Figure:** System of towns, after refinement $j = 3$
Results: $\psi$

Figure: Function $\psi$ generated after 11 iterations.
Table of Contents

1 Motivation

2 Characteristics of Inner Functions

3 Construction of a Lipschitz Inner Function

4 Implementation and Results

5 Moving Forward
Given $f$, need to construct outer function:

- Dependent on $f$
- Only need one (Lorenz 1966)
- If $f$ is differentiable, needs to cancel out the non-differentiability of inner function
- Constructed iteratively by bounding oscillation and refining
- Implement in code...