

Formalizing a Lipschitz continuous version of Kolmogorov's
superposition theorem

Ben Goodrich

March 25, 2025

Chapter 1

Introduction

This project is not yet complete, but intends to formalize a strengthening of the following theorem. Let $\mathbb{I} = [0, 1]$ and let $n \geq 2$ be an given integer. We use C^d to indicate the space of functions defined on a domain whose d -th derivative is continuous and use \circ to indicate function composition.

Theorem (Kolmogorov 1957). For every $f \in C(\mathbb{I}^n) \mapsto \mathbb{R}$, there exist increasing $\psi_{p,q} \in C(\mathbb{I}) \mapsto \mathbb{R}$ that do not depend on f and generally non-monotone $\chi_q \in C(\mathbb{R}) \mapsto \mathbb{R}$ that depend on f and $\psi_{p,q}$, such that $f(x_0, x_1, \dots, x_{n-1}) = \sum_{q=0}^{2n} \chi_q \circ \Psi_q(x_0, x_1, \dots, x_{n-1})$, where $\Psi_q(x_0, x_1, \dots, x_{n-1}) \equiv \sum_{p=0}^{n-1} \psi_{p,q}(x_p)$.

This theorem is called the Kolmogorov Superposition Representation (KSR) or the Kolmogorov Superposition Theorem (KST) because Kolmogorov proved that a given multivariable “target” function, f , could be represented as a finite sum of compositions of univariate functions.

The relevance of the KRT continues to be debated in the neural networks literature. On one side, many want to utilize the KSR to approximate unknown or expensive target functions. However, the “inner” functions, $\psi_{p,q}$, and “outer” functions, χ_q , are not elementary, in the sense that they cannot be computed with a finite number of calls to basic functions. Thus, the KSR requires harmonic analysis and other side is skeptical that the it will ever be of much practical use.

Although the KST is perhaps surprising in its generality and simplicity, it has been proven in several ways, raising the question of what value is added by an additional proof? The extant proofs of the KST are of three kinds (and none have been formalized):

- **Nonconstructive:** The most elegant proofs utilize the Baire Category Theorem to prove the *existence* of inner and outer functions that satisfy the KST but do not suggest how to construct them.
- **Constructive:** Other proofs are considered to be constructive but do not even suggest a feasible algorithm to implement them in floating-point arithmetic.
- **Constructed:** Two constructions have been implemented in software but both utilize bad inner functions and worse outer functions, making them unworkable in floating-point arithmetic.

By “bad” inner functions, we specifically mean they are in the Devil’s Staircase family, that is, their derivative is zero almost everywhere on \mathbb{I} but does not exist on a dense set of points. By “worse” outer functions, we specifically mean that they are nowhere differentiable. Thus, even if f were differentiable at some point, the chain rule does not apply to a compositions involving nowhere differentiable functions, so it cannot be used to compute a gradient under the original KSR. Most algorithms in statistics and supervised learning rely on gradients as part of some larger algorithm, so the KSR has not been useful for those purposes.

Our goal is to construct decent inner functions and adequate outer functions that satisfy the KST and allow the chain rule to compute the gradient of f , which can then be implemented to floating-point arithmetic. Vitushkin and Khenkin proved that the inner functions in the KST cannot all be continuously differentiable. Thus, by “decent” inner functions, we specifically mean Lipschitz continuous functions, which satisfy $|\psi_{p,q}(x) - \psi_{p,q}(x')| \leq K|x - x'|$ for some positive but finite constant K . The KST literature often refers to a more general definition of Lipschitz continuity (which is also termed Hölder continuity) where $|\psi_{p,q}(x) - \psi_{p,q}(x')| \leq K|x - x'|^\alpha$. Fridman and Kahane proved that the inner functions can all be weak contractions, that is, $\alpha = 1$ and $K = 1$.

Sprecher proved that there are outer functions that satisfy the KST for any target function when $\psi_{p,q}(x_p) \equiv \lambda^p \varphi(x_p - \frac{q}{2^n})$, where $\lambda > 0$ is an irrational number and φ is a weak contraction that we call the “core” function. However, the only known attempt to implement the Fridman-Sprecher approach in software was made by Actor, which was unsuccessful.

Actor did succeed in reparameterizing the Hölder continuous core function proposed by Sprecher, corrected by Köppen, and proven correct by Braun and Griebel. Actor’s function is parameterized via its arc length and is Lipschitz continuous by construction. While there is graphical evidence that strongly suggests Actor’s function is compatible with the KST, some aspects of the approach are not fully proved, and it requires an expensive numerical inversion of Braun and Griebel’s function just to achieve a few decimal places of accuracy.

Montanelli and Yang notes the importance of constructing a Lipschitz continuous KST

We have proven upper bounds for the approximation of multivariate functions $f : [0, 1]^n \mapsto \mathbb{R}$ by deep ReLU networks, for which the curse of dimensionality is lessened. The depth and the size of the networks to approximate such functions f grow like $\mathcal{O}(\varepsilon^{-\log n})$, as opposed to $\mathcal{O}(\varepsilon^{-n})$. The proof is based on the ability of very deep ReLU networks to implement the Kolmogorov-Arnold superposition theorem.

There are many ways in which this work could be fruitfully continued. If we were able to construct a Lipschitz continuous inner function, we would be able to obtain $\mathcal{O}(\varepsilon^{-1})$ estimates. Actor and Knepley designed in 2017 an algorithm to compute a Lipschitz continuous inner function, but they did not provide a method to compute the outer functions.

This quote reflects the fact that the outer functions depend on the inner functions, in addition to f . The smoother are the inner functions, the easier it is to obtain “adequate” outer functions that approximate a target function to a given degree of accuracy. In short, a Lipschitz continuous KST would overcome (to first order) the curse of dimensionality when approximating y on \mathbb{I}^n .

Also, target functions in statistics, such as log-likelihood functions, typically depend on N data points and the cost of evaluating them directly is polynomial (at best) in N . Once the outer functions have been calibrated to the target function, the KSR of the target function can be evaluated in parallel in essentially constant wall time regardless of n or N . These savings could be

enormous in computationally intensive applications, such as Markov Chain Monte Carlo or Large Language Models.

There are several open questions that are not even investigated in the KST literature:

1. Can some proper subset of the inner functions be continuously differentiable (or smoother)? We show that if $0 < q < 2n$, then $\psi_{p,q}$ can be twice continuously differentiable.
2. Does the core function depend on n , apart from being evaluated at $x - \frac{q}{2n}$? We show that φ can be universal, although λ does depend on n in a simple fashion.
3. Can the set of points where φ is not differentiable be finite? We show that φ is not differentiable at 1 and -1 but is differentiable on $(-1, 1)$.
4. Can φ be approximated to floating-point precision? We show that φ can be quickly approximated to any degree of accuracy.

Our approach is unique in that we first construct a function with the requisite properties and then show that it is a viable core function in the KSR. Chapter 2 uses basic real analysis — most of which is already formalized in the Mathlib library used by the Lean software — to show that our candidate for ψ is increasing, Lipschitz continuous, and not continuously differentiable. In addition, since φ can be represented as a series, it is essential to demonstrate that the series converges uniformly, which is an improvement over the other constructions in the KST literature whose functions only converge pointwise. As a result, φ can be evaluated to double precision in fewer than a hundred floating point operations, and its derivative can be evaluated at double that cost (albeit to less accuracy).

Chapter 3 establishes in Lean that $\Psi_q(x_0, x_1, \dots, x_{n-1}) \equiv \sum_{p=1}^n \lambda^{p-1} \varphi(x_p - \frac{q}{2n})$ is injective for each q when λ is a transcendental number. The proof is only slightly different from the one Sprecher has relied upon for decades that essentially extended the rational number field with an irrational λ . Much of this can leverage the theorems on cyclotomic fields that are already in Mathlib.

Chapter 4 is the least complete but is devoted to the outer functions in the KSR. Kahane utilized the notion of a Helson set, which is a subset with the characterizing property that every continuous function over this subset can be represented as a convergent Fourier series. Kahane proved necessary and sufficient conditions for a subset of \mathbb{T}^{2n+1} to be a Helson set.

Chapter 2

Core Function

2.1 Main Definitions

The atomic element of our KSR is the first-kind Chebyshev polynomial. Let j be a generic integer that is usually non-negative, but is used in various ways in different contexts in this report.

Definition 1 (Chebyshev polynomial of the first kind). Let r be a real number. The first-kind Chebyshev polynomial in r with index j is defined via a three-term recursion as

$$T_j(r) \equiv \begin{cases} 1 & \text{if } j = 0 \\ r & \text{if } j = 1 \\ 2rT_{j-1}(r) - T_{j-2}(r) & \text{if } j > 1. \end{cases}$$

Remark. The Mathlib library used by Lean (version 4) actually defines T_j as a function of j (which can be negative) that returns a polynomial (with coefficients in any commutative ring). Such generality is unnecessary here, but Mathlib also provides a metatheorem that can be used to tersely prove many properties of Chebyshev polynomials via double induction on j . Our notation, $T_j(r)$, corresponds to simply evaluating such a polynomial at a symbolic real number.

We specialize $j = 2^m$ to define partial sums that approach our candidate for a core function.

Definition 2 (k -th order approximator). $\varphi_k(r) \equiv \frac{3}{7} + \frac{8^{-k}}{14} + \frac{1}{2} \sum_{m=0}^k 8^{-m} T_{2^m}(r)$, where $k \in \mathbb{N}$.

Definition 3 (Core function). $\varphi(r) \equiv \frac{3}{7} + \frac{1}{2} \sum_{m=0}^{\infty} 8^{-m} T_{2^m}(r)$.

For an overview of expanding a function in terms of Chebyshev polynomials, see Trefethen.

Definition 4 (Normalizer). $\Lambda \equiv \sum_{p=0}^{n-1} \lambda^p$, where $\lambda > 0$ and $n \geq 2$.

Definition 5 (Scale factor). $\lambda_p \equiv \frac{\lambda^p}{\Lambda} \in (0, 1)$.

Definition 6 (Inner function). $\psi_{p,q}(x) \equiv \lambda_p \varphi(x - \frac{q}{2n})$, where $x \in \mathbb{I}$ and $q \in \{0, 1, \dots, 2n\}$.

2.2 Repeatedly Used Lemmas

Our Chebyshev polynomials also follow a *two-term* recursion that drives nearly all other lemmas.

Lemma A (Two-term recursion). $T_{m+1} = 2T_m^2 - 1$, where $m \in \mathbb{N}$.

Proof. Since $2^{m+1} = 2 \times 2^m$, we can write $T_{m+1} = T_{2 \times 2^m}$. It is already proven by induction in Mathlib (TODO: add succinct explanation) that a Chebyshev polynomial whose index is a product of integers can be written as a composition of two Chebyshev polynomials, so $T_{2 \times 2^m} = T_2 \circ T_m$. It follows from Definition 1 that $T_2(r) = 2r^2 - 1$, so $T_2 \circ T_m = 2T_m^2 - 1 = T_{m+1}$. \square

Lemma B (Bounds). $T_{2^m}(r)^2 \leq 1$ if and only if $r^2 \leq 1$.

Proof. Since $T_{2^0}(r) = r$, having $r^2 \leq 1$ is necessary and sufficient when $m = 0$. Assume it holds for m , so that we can use Lemma A to show it then holds for $T_{2^{m+1}}(r)^2 = (2T_{2^m}(r)^2 - 1)^2$. Under the inductive hypothesis that $T_{2^m}(r)^2 \leq 1$, then the entire right-hand side is less than one in magnitude. Conversely, if $T_{2^m}(r)^2 > 1$, then the right-hand side remains greater than one in magnitude. \square

Lemma C (Fixed points). *The only two fixed points of $2T_{2^m}(r)^2 - 1 \mapsto T_{2^{m+1}}(r)$ are $-\frac{1}{2}$ and 1.*

Proof. If $T_{2^{m+1}}(r) = T_{2^m}(r)$ under Lemma A, then a quadratic equation, $2T_{2^m}(r)^2 - T_{2^m}(r) - 1 = (2T_{2^m}(r) + 1)(T_{2^m}(r) - 1) = 0$ whose only two solutions are $-\frac{1}{2}$ and 1 defines the fixed points. \square

Remark. Both fixed points are repellent because the derivative of the mapping at a fixed point greater in magnitude than 1. However, this has not been formalized yet since it is not a priority.

If either fixed point is reached, the remaining terms in φ_k become a partial geometric series.

Lemma D (Partial geometric series). *If $b > 1$, then $\sum_{m=j}^k b^{-m} = \frac{b^{1-j} - b^{-k}}{b-1}$.*

Proof. $\sum_{m=j}^k b^{-m} = \sum_{m=j}^k \left(\frac{1}{b}\right)^m \equiv S$, which is in the usual form of a geometric series with a ratio of $\frac{1}{b}$ whose partial sums are already proven in Mathlib. The usual method of proof is to note that since $S - \frac{1}{b}S = \sum_{m=j}^k \left[\left(\frac{1}{b}\right)^m - \left(\frac{1}{b}\right)^{m+1}\right] = \left(\frac{1}{b}\right)^j - \left(\frac{1}{b}\right)^{k+1}$, then $S = \frac{\left(\frac{1}{b}\right)^j - \left(\frac{1}{b}\right)^{k+1}}{1 - \frac{1}{b}} = \frac{b^{-j} - b^{-k-1}}{\frac{b-1}{b}} = \frac{b^{1-j} - b^{-k}}{b-1}$. \square

Lemma E (Symmetry). $T_{2^m}(-r) = (-1)^{2^m} T_{2^m}(r)$.

Proof. This also holds for Chebyshev polynomials whose index is not a power of two, but the general case is not yet proven in Mathlib, so we will prove it for $j = 2^m$. If $m = 0$, then $T_{2^0}(r) = r$ so the claimed equation holds. Per the two-term recursion in Lemma A, we need to show that $2T_{2^m}(-r)^2 - 1 = T_{2^{m+1}}(-r) = T_{2^{m+1}}(r) = 2T_{2^m}(r)^2 - 1$ because $(-1)^{2^m} = 1$ when $m > 0$. Substituting for $T_{2^m}(-r)$ with the inductive hypothesis implies $2 \left[(-1)^{2^m} T_{2^m}(r)\right]^2 - 1 = 2T_{2^m}(r)^2 - 1$. \square

Lemma F (Functional equation). $\varphi_k(-r) = \varphi_k(r) - r$.

Proof. Per Lemma E, the only term that differs in Definition 3 between the r and $-r$ cases is when $m = 0$ and $\frac{1}{2}8^0 T_{2^0}(\mp r) = \frac{\mp r}{2}$. Thus, $\varphi_k(r) - \varphi_k(-r) = \frac{r}{2} - \frac{-r}{2} = r$. \square

Lemma G (Linearly shifted sum-of-squares). $\varphi_k(r) = \frac{5}{14} + \frac{8^{-k}}{7} + \frac{r}{2} + \frac{1}{8} \sum_{m=0}^{k-1} 8^{-m} T_{2^m}(r)^2$.

Proof. Pull the $T_{2^0}(r) = r$ term out of the series in Definition 2 and then shift the index of the summation while applying the two-term recursion from Lemma A to all the remaining terms to obtain $\varphi_k(r) = \frac{3}{7} + \frac{8^{-k}}{14} + \frac{r}{2} + \frac{1}{2} \sum_{m=0}^{k-1} 8^{-(m+1)} (2T_{2^m}(r)^2 - 1)$. We can pull a factor of $\frac{1}{8}$ out of the sum and then separate it to obtain $\varphi_k(r) = \frac{3}{7} + \frac{8^{-k}}{14} + \frac{r}{2} + \frac{1}{8} \sum_{m=0}^{k-1} 8^{-m} T_{2^m}(r)^2 - \frac{1}{16} \sum_{m=0}^{k-1} 8^{-m}$. Apply Lemma D with $b = 8$ to the last term and simplify the constant to obtain the claimed result. \square

2.3 Convergence

Lemma 2.3.1. *Let $\mathbb{T} = [-1, 1]$. If $t \in \mathbb{T}$, then the sequence of partial sums, $\{\varphi_k(t)\}$ forms a uniformly Cauchy sequence, but if $r \notin \mathbb{T}$, then the infinite sum in $\varphi(r)$ diverges.*

Proof. To prove the if direction, we utilize the Weierstrass M-test, which is sufficient to prove that the sequence of partial sums forms a uniformly Cauchy sequence. Lemma B implies $T_{2^m}(t) \in \mathbb{T}$. Since $|8^{-m} T_{2^m}(t)| \leq 8^{-m}$ and $\frac{1}{2} \sum_{m=0}^{\infty} 8^{-m} = \frac{4}{7}$ by Lemma D with $b = 8$, φ is well-defined on \mathbb{T} .

To prove the only if direction, we utilize the last term test. It suffices to show that if $|r| > 1$, then the limit of $8^{-m} T_{2^m}(r)$ as $m \uparrow \infty$ is not zero. When $r > 1$, it is already proven in Mathlib (TODO: add succinct explanation) that $T_j(r) = \cosh(j\theta)$ where θ is implicitly defined by $\cosh \theta = r$. Mathlib unfortunately lacks the \cosh^{-1} function but has \sinh^{-1} , so we can explicitly define $\theta \equiv \sinh^{-1}(r^2 - 1) > 0$. Mathlib does prove that $\cosh(\sinh^{-1}(r^2 - 1)) = \sqrt{1 + r^2 - 1} = r > 1$. Moreover, $\cosh(2^m \theta) = \frac{e^{2^m \theta} + e^{-2^m \theta}}{2}$, so $\ln \frac{T_{2^m}(r)}{2^{3m}} = 2^m \theta + \ln(1 + e^{-2^{m+1} \theta}) - (3m + 1) \ln 2$. As $m \uparrow \infty$, a Maclaurin-series expansion of $\ln(1 + e^{-2^{m+1} \theta})$ reveals it approaches the negligibly small value of $e^{-2^{m+1} \theta}$, and since $2^m \theta > 0$, it approaches ∞ as $m \uparrow \infty$ at a much faster rate than $-(3m + 1) \ln 2$ approaches $-\infty$. Thus, there is no limit of $8^{-m} T_{2^m}(r)$ as $m \uparrow \infty$ when $r > 1$. When $r < -1$, Lemma E shows that the only term in the infinite sum that changes is when $m = 0$ and $T_{2^0}(\mp r) = \mp r$. Thus, the infinite sum also diverges when $r < -1$. \square

Remark. Since the error in approximating $\varphi(t)$ with $\varphi_k(t)$ is less than $\frac{2^{-3k}}{14}$, the approximation error if $k = 17$ is less than $\frac{2^{-52}}{7}$, where 2^{-52} is the smallest double-precision number that can be added to 1 without underflow. Thus, only 18 iterations of an unrolled loop with only a few floating-point operations per iteration are needed to recursively evaluate $\varphi_{17}(t)$ from Lemma G.

Remark. If $t = \frac{a}{b}$ where $a \in \mathbb{Z}$ and $b \in \mathbb{N}$ such that $|a| < b$ and a has no common factor with b , then $\varphi_k(\frac{a}{b})$ is also rational because the only operations in Definition 3 are arithmetic. However, $\varphi(\frac{a}{b})$ is generally irrational. The only five known rational values of φ occur when one of the fixed points from Lemma C is quickly reached: $\varphi(1) = 1$ (implicit in Lemma 2.3.1), $\varphi(-1) = \varphi(1) - 1 = 0$ (from Lemma F), $\varphi(-\frac{1}{2}) = \frac{3}{7} + \frac{1}{2} \sum_{m=0}^{\infty} 8^{-m} (-\frac{1}{2}) = \frac{1}{7}$ (from Lemmas C and D), $\varphi(\frac{1}{2}) = \varphi(-\frac{1}{2}) + \frac{1}{2} = \frac{9}{14}$ (from Lemma F), and $\varphi(0) = \frac{3}{7} + \frac{0}{2} - \frac{1}{8} + \frac{1}{2} \sum_{m=0}^{\infty} 8^{-m} = \frac{3}{8}$ (from Lemmas C and D). These five

point coincide with the only five values where cosine and sine are rational according to Niven's theorem. Although the partial sums converge uniformly with an error bound proportional $\frac{1}{8k}$, they do so with denominators that are too large for $\varphi(\frac{a}{b})$ to be a Liouville number, so they may well be algebraic rather than transcendental. In addition, if $T_{2^m}(\frac{a}{b})$ is never either of the fixed points from Lemma C, then the numerators of the partial sums do not have a limit.

Lemma 2.3.2. *φ is not differentiable at -1 and 1 , which are essential logarithmic singularities.*

Proof. The derivative is defined as $\dot{\varphi}(t) \equiv \lim_{h \rightarrow 0} \frac{\varphi(t+h) - \varphi(t)}{h}$. Per Lemma 2.3.1, if $t+h \notin \mathbb{T}$, then this limit is undefined, which can occur if both $t = 1$ and $h > 0$ or if both $t = -1$ and $h < 0$. In other words, φ is merely right differentiable at $t = 1$ and left differentiable at $t = -1$. \square

Proposition 2.3.3. *If $q = 0$ or $q = 2n$, then $\psi_{p,q} \notin C^1(\mathbb{I})$.*

Proof. Recall from Definition 6 that $\psi_{p,q}(x) \equiv \lambda_p \varphi(x - \frac{q}{2n})$. Per Lemma 2.3.2, if $x+h - \frac{q}{2n} \notin \mathbb{T}$, then $\dot{\varphi}$ does not exist at that point and hence neither does $\dot{\psi}_{p,q}$ exist at that point. If $h > 0$, that exceptional case can occur if and only if both $x = 1$ and $q = 0$, and if $h < 0$, it can also occur if and only if both $x = 0$ and $q = 2n$. \square

Remark. Thus, our candidate for a core function narrowly avoids the fact proven by Vituskin and Khenkin that the KST would not hold if *all* the inner functions were continuously differentiable.

Lemma 2.3.4. *$\varphi \in C(\mathbb{T})$.*

Proof. φ_k in Definition 2 is a sum of polynomials and thus is continuous. Per Lemma 2.3.1, $\{\varphi_k\}$ converges uniformly to φ on \mathbb{T} . The uniform limit of a sequence of continuous functions is continuous, so φ is continuous over \mathbb{T} but no smoother in light of Lemma 2.3.2. \square

Lemma 2.3.5. *If $q = 0$ or $q = 2n$, then $\psi_{p,q} \in C(\mathbb{I})$.*

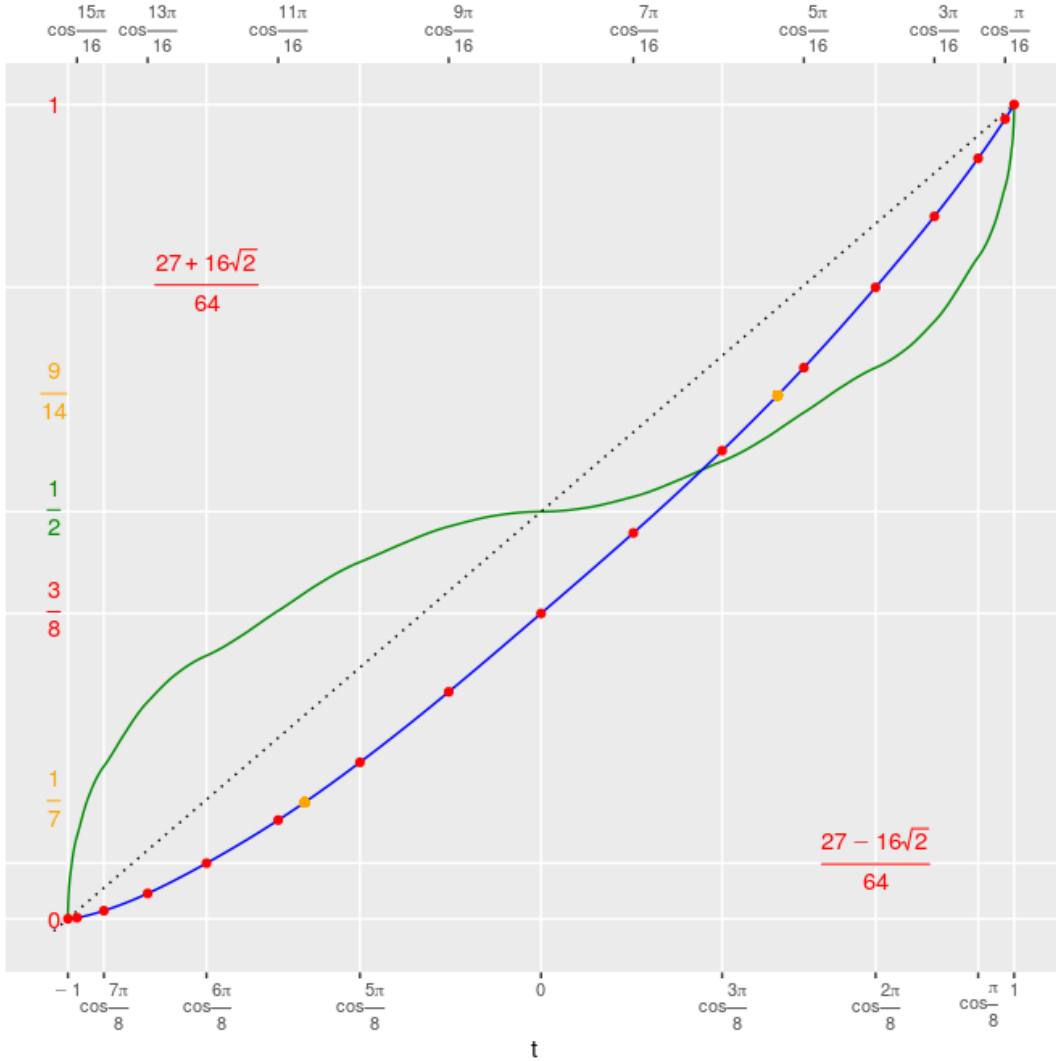
Proof. Recall from Definition 6 that $\psi_{p,q}(x) \equiv \lambda_p \varphi(x - \frac{q}{2n})$, so Lemma 2.3.4 in fact implies $\psi_{p,q} \in C(\mathbb{I})$ for any p and q , but we will strengthen this for $0 < q < 2n$ below in Proposition 2.4.10. \square

However, these last few results lack nuance because the derivative is undefined only at one or both endpoints of a domain, and the functions are quite smooth on the interior, as is shown in Figure 2.1. Both φ_{17} (in blue) and its derivative (in green) are clearly increasing functions from \mathbb{T} to \mathbb{I} . It appears that a small perturbation to φ would violate one or more of our three requirements for a core function in the KSR (that are proven in Section 2.5): increasing over \mathbb{T} , not continuously differentiable over \mathbb{T} , and having a bounded derivative wherever it exists on \mathbb{T} . There may well be other ways to define a core function that has these three properties, but no one really has. Moreover, it seems difficult to construct one that converges more rapidly than φ_k does.

2.4 Differentiability

The fact that $\{\varphi_k\}$ converges uniformly to φ per Lemma 2.3.1 does not, by itself, imply that $\{\dot{\varphi}_k\}$ converges in any sense to $\dot{\varphi}$, in part because the derivative of φ does not even exist at -1 and 1 per Lemma 2.3.2. The purpose of this subsection is to establish rigorously that the intuition from the green line in Figure 2.1 holds on the open interval, $(-1, 1)$, due to the uniform convergence of the derivatives, which can also be established using the Weierstrass M-test. Since T_{2^m} and φ_k are polynomials, their derivatives are also polynomials, although their limit as $k \uparrow \infty$ is not.

Figure 2.1: The (well-approximated) core function and its derivative



The blue line is for φ_{17} , which approximates φ to double-precision. φ_{17} is a polynomial of degree $2^{17} = 131,072$ whose exponents are all even numbers except for the linear term, which oscillates around a line almost 200 times per pixel with amplitudes of about a quadrillionth. The green line is for the exact derivative of φ_{17} , which approximates the derivative of φ on $(-1, 1)$ in which case the latter exists, as is proven later in Lemma 2.5.1. The black dotted line is for φ_0 . The red points are discussed in Chapter 3 and are $\varphi(\tau)$ at Chebyshev-Lobatto nodes when $\tau = -\cos\frac{j\pi}{16}$ and j is an integer between 0 and 16 inclusive. The two orange points are $\varphi(-\frac{1}{2}) = \frac{1}{7}$ and $\varphi(\frac{1}{2}) = \frac{9}{14}$, which along with the nodal values $\varphi(-1) = 0$, $\varphi(0) = \frac{3}{8}$, and $\varphi(1) = 1$ are the only five known rational values of φ that are discussed in the second remark following Lemma 2.3.1.

Definition 7 (Chebyshev polynomial of the second kind). The second-kind Chebyshev polynomial in r with index $j \geq -1$ is defined recursively as

$$U_j(r) \equiv \begin{cases} 0 & \text{if } j = -1 \\ 1 & \text{if } j = 0 \\ 2r & \text{if } j = 1 \\ 2rU_{j-1}(r) - U_{j-2}(r) & \text{if } j > 1 \end{cases}$$

Lemma 2.4.1. *If $j \geq 0$, then $\dot{T}_j(r) = jU_{j-1}(r)$.*

Proof. This is already proven by induction in Mathlib. TODO: Add a succinct explanation. \square

Lemma 2.4.2. $U_{-1+2^{m+1}}(r) = 2T_{2^m}(r)U_{-1+2^m}(r)$.

Proof. Differentiate both sides of the two-term recursion from Lemma A and utilize Lemma 2.4.1 to obtain $2^{m+1}U_{-1+2^{m+1}}(r) = 2^2T_{2^m}(r)2^mU_{-1+2^m}(r)$, and then divide both sides by 2^{m+1} . \square

Lemma 2.4.3. $\dot{\varphi}_k(r) = \frac{1}{2} \sum_{m=0}^k 4^{-m}U_{-1+2^m}(r)$.

Proof. Applying Lemma 2.4.1 to each non-constant term in Definition 2 yields this result. \square

Lemma 2.4.4. *If $t \in \mathbb{T}$, then the sequence of partial sums, $\{\dot{\varphi}_k(t)\}$ forms a uniformly Cauchy sequence.*

Proof. We utilize the Weierstrass M-test, which is sufficient to prove that the sequence of partial sums forms a uniformly Cauchy sequence. If $m = 0$, then $U_0(t) = t$ has a minimum of -1 and a maximum of 1 . Per Lemma B, if $t \in \mathbb{T}$, then $T_{2^m}(t) \in \mathbb{T}$, so Lemma 2.4.2 indicates that the range of $U_{-1+2^{m+1}}$ doubles with each m . Thus, $U_{-1+2^{m+1}}$ has a minimum of -2^m and a maximum of 2^m . Since $|4^{-m}U_{-1+2^m}| \leq 2^{-m}$ and $\frac{1}{2} \sum_{m=0}^{\infty} 2^m = 1$ per Lemma D with $b = 2$, as $k \uparrow \infty$, $\{\dot{\varphi}_k(t)\}$ converges uniformly on \mathbb{T} . \square

Lemma 2.4.5. $\lim_{k \uparrow \infty} \dot{\varphi}_k$ is a weak derivative of φ that coincides with $\dot{\varphi}$ on $(-1, 1)$.

Proof. The uniform convergence of $\{\dot{\varphi}_k(t)\}$ in Lemma 2.4.4 justifies interchanging the limits in

$$\dot{\varphi}_\infty(t) = \lim_{h \rightarrow 0} \lim_{k \uparrow \infty} \frac{[\varphi_k(t+h) - \varphi_k(t)]}{h} = \lim_{k \uparrow \infty} \lim_{h \rightarrow 0} \frac{\varphi_k(t+h) - \varphi_k(t)}{h} = \frac{1}{2} \sum_{m=0}^{\infty} 4^{-m}U_{-1+2^m}(t),$$

where the last line follows from Lemma 2.4.3. Per Lemma 2.3.1, since $\{\varphi_k(t)\}$ converges to $\varphi(t)$ for at least one (actually all) value(s) of $t \in \mathbb{T}$, $\dot{\varphi}(t)$ exists on any proper *subinterval* of \mathbb{T} . \square

Lemma 2.4.6. $(1 - r^2)\ddot{T}_j(r) - t\dot{T}_j(r) + j^2T_j(r) = 0$.

Proof. Although this lemma has been known for decades, it has not been formalized in Mathlib, so we follow the proof in Ponton. From Definition 1, if $j = 0$, then $T_0(r) = 1$, so $\dot{T}_0(r) = 0$ and $\ddot{T}_0(r) = 0$, which are consistent with the above Ordinary Differential Equation (ODE). Similarly, if $j = 1$, then $T_1(r) = t$ so $\dot{T}_1(r) = 1$ and $\ddot{T}_1(r) = 0$, which are also consistent with the above ODE.

If $j > 1$, we can differentiate the three-term recursion, $T_j(r) = 2rT_{j-1}(r) - T_{j-2}(r)$, to obtain

$$\begin{aligned}\dot{T}_j(r) &= 2T_{j-1}(r) + 2r\dot{T}_{j-1}(r) - \dot{T}_{j-2}(r) \\ \ddot{T}_j(r) &= 4\dot{T}_{j-1}(r) + 2r\ddot{T}_{j-1}(r) - \ddot{T}_{j-2}(r)\end{aligned}$$

If we substitute these three relations into the purported ODE and factor, we need to show that

$$\begin{aligned}0 &= 2r \left[(1-r^2)\ddot{T}_{j-1}(r) - r\dot{T}_{j-1}(r) + (j-1)^2 T_{j-1}(r) \right] \\ &\quad - \left[(1-r^2)\ddot{T}_{j-2}(r) - r\dot{T}_{j-2}(r) + (j-2)^2 T_{j-2}(r) \right] \\ &\quad + 4(j-1) \left[(1-r^2)U_{j-2}(r) + rT_{j-1}(r) - T_{j-2}(r) \right].\end{aligned}$$

In each of the first two lines, the term in brackets is zero due to the inductive hypothesis, so we need to show that the term in brackets in the third line is also zero. It is already proven by induction in Mathlib (TODO: add succinct explanation) that $(1-r^2)U_{j-2}(r) = rT_{j-1}(r) - T_{j-2}(r)$, so substituting and then expressing $T_j(r)$ via the three-term recursion yields the claimed result. \square

Lemma 2.4.7. *If $t \in (-1, 1)$, then $\ddot{\varphi}_k(t) = \frac{t}{1-t^2}\dot{\varphi}_k(t) - \frac{1}{2(1-t^2)} \sum_{m=0}^k 2^{-m}T_{2^m}(t)$.*

Proof. Differentiate Definition 2 twice to obtain $\ddot{\varphi}_k(t) = \frac{1}{2} \sum_{m=0}^k 8^{-m}\ddot{T}_{2^m}(t)$. Substitute inside the sum using Lemma 2.4.6 after isolating $\ddot{T}_{2^m}(t)$, and then simplify using Lemma 2.4.3. \square

Remark. If $t^2 = 1$, then $\ddot{\varphi}_k(\mp 1) = \frac{2^{k+1}+2^{-k}-3}{6}$ by L'Hospital's rule, but this edge case is irrelevant.

Lemma 2.4.8. *If $t \in (-1, 1)$, then the sequence of partial sums, $\{\ddot{\varphi}_k(t)\}$ forms a uniformly Cauchy sequence.*

Proof. Lemma 2.4.4 establishes that $\dot{\varphi}_k(t)$ in Lemma 2.4.7 converges uniformly on \mathbb{T} . The same proof of the “if” direction of Lemma 2.3.1 but now with $b = 2$ shows that the other sum also converges uniformly. \square

Lemma 2.4.9. $\lim_{k \uparrow \infty} \ddot{\varphi}_k$ is a weak second derivative of φ that coincides with $\ddot{\varphi}$ on $(-1, 1)$.

Proof. The uniform convergence of $\{\ddot{\varphi}_k(t)\}$ from Lemma 2.4.8 justifies interchanging the limits in

$$\begin{aligned}\ddot{\varphi}_\infty(t) &= \lim_{h \rightarrow 0} \lim_{k \uparrow \infty} \frac{[\dot{\varphi}_k(t+h) - \dot{\varphi}_k(t)]}{h} = \lim_{k \uparrow \infty} \lim_{h \rightarrow 0} \frac{\dot{\varphi}_k(t+h) - \dot{\varphi}_k(t)}{h} = \lim_{k \uparrow \infty} \ddot{\varphi}_k(t) \\ &= \frac{t}{1-t^2}\dot{\varphi}_\infty(t) + \frac{1}{2(1-t^2)} \sum_{m=0}^{\infty} 2^{-m}T_{2^m}(t).\end{aligned}$$

Since $\dot{\varphi}$ does not exist at -1 or 1 , neither does $\ddot{\varphi}$ exist at those two points. However, per Lemma 2.4.4, since $\{\dot{\varphi}_k(t)\}$ converges to $\dot{\varphi}(t)$ for at least one (actually all) value(s) of $t \in (-1, 1)$, $\ddot{\varphi}_\infty$ coincides with $\ddot{\varphi}(t)$ on $(-1, 1)$ and evaluates to the extended real number, ∞ , if $t^2 = 1$. \square

Proposition 2.4.10. *If $0 < q < 2n$, then $\psi_{p,q} \in C^2(\mathbb{I})$.*

Proof. Recall from Definition 6 that $\psi_{p,q}(x) \equiv \lambda_p \varphi(x - \frac{q}{2n})$. If $0 < q < 2n$, then $x - \frac{q}{2n}$ cannot be -1 or 1 , in which case Lemma 2.4.8 implies that $\psi_{p,q}$ is twice differentiable. \square

Remark. Proposition 2.4.10 does not contravene the conclusion of Vitushkin and Khenkin that the KST cannot hold if *all* the inner functions were continuously differentiable because Proposition 2.3.3 proves that if $q = 0$ or $q = 2n$, then $\psi_{p,q} \notin C^1(\mathbb{I})$. Technicalities aside, our construction is much smoother than any other one in the KSR literature.

Remark. It seems implausible for φ to have a third derivative over any proper subinterval of \mathbb{T} because differentiating once more under the sum in Lemma 2.4.8 would yield a term, $\sum_{m=0}^{\infty} U_{-1+2^m}(t)$, that diverges unless t is a root of some U_{-1+2^m} and thus a root of $U_{-1+2^{m+1}}$ per Lemma 2.4.2.

Remark. $\dot{\varphi}_{17}(t)$ and $\ddot{\varphi}_{17}(t)$ are less accurate near the endpoints of \mathbb{T} than is $\varphi_{17}(t)$ due to the slower rates of convergence in the geometric series. But most numerical applications do not require extreme accuracy in the derivatives.

2.5 Requisite Properties

This section merely proves what is already strongly suggested by Figure 2.1, namely that φ is weakly convex, strictly increasing, and a weak contraction.

Lemma 2.5.1. $\dot{\varphi}_k(-r) = 1 - \dot{\varphi}_k(r)$.

Proof. Differentiate both sides of the functional equation in Lemma F. \square

Lemma 2.5.2. $\ddot{\varphi}_k(-r) = \ddot{\varphi}_k(r)$.

Proof. Differentiate both sides of the functional equation in Lemma 2.5.1. \square

Lemma 2.5.3. *If $k > 0$, then φ_k is a strictly convex function over \mathbb{T} , and if $k = 0$, then φ_0 is weakly convex over \mathbb{T} .*

Proof. A twice-differentiable function, such as φ_k , is strictly convex on \mathbb{T} if and only if its second derivative is positive on \mathbb{T} . Suppose on the contrary that $\ddot{\varphi}_k(t) = 0$ for some $t \in \mathbb{T}$. Since $\ddot{\varphi}_k$ is an even function per Lemma 2.5.2, $\ddot{\varphi}_k(-t) = 0$ as well. Subtracting the latter equation from the former using the expression for the second derivative in Lemma 2.4.7 implies $\frac{t}{1-t^2} \dot{\varphi}_k(t) - \frac{-t}{1-t^2} \dot{\varphi}_k(-t) = \frac{1}{2(1-t^2)} \sum_{m=0}^k 2^{-m} (T_{2^m}(t) - T_{2^m}(-t))$. The left-hand side simplifies to $\frac{t}{1-t^2}$ under Lemma 2.5.1, while the right-hand side also simplifies to $\frac{t}{1-t^2}$ under Lemma E. Together that implies that $\ddot{\varphi}_k$ would be a horizontal line at zero, which is actually the case when $k = 0$ and $\varphi_0(t) = \frac{1+t}{2}$ is a line segment. If $k > 0$ and $t = 0$, then Lemma 2.4.7 implies $\ddot{\varphi}_k(0) = \frac{0}{1} \dot{\varphi}_k(0) - \frac{1}{2} \left(0 - \frac{1}{2} + \sum_{m=2}^k 2^{-m} \right) = 2^{-k-1} > 0$, where the last equality follows from Lemma D with $b = 2$. This example contradicts the premise that $\ddot{\varphi}_k$ is a horizontal line at zero when $k > 0$ and certifies that the second derivative is strictly positive over \mathbb{T} rather than strictly negative. \square

Lemma 2.5.4. φ is a weakly convex function over \mathbb{T} .

Proof. Per Lemma 2.5.3, φ_k is a strictly convex when $k > 0$, and per Lemma 2.3.1 $\{\varphi_k\}$ converges uniformly to φ . The uniform limit of a sequence of strictly convex functions is at least a weakly convex function. Although $\lim_{k \uparrow \infty} 2^{-k-1} = 0$ and $\ddot{\varphi}(0) = 0$ is an inflection point, $\ddot{\varphi}$ can never be negative on \mathbb{T} , so φ is weakly convex. \square

Lemma 2.5.5. *If $t \in \mathbb{T}$, then $2^{-k-1} \leq \dot{\varphi}_k(t) \leq 1 - 2^{-k-1}$.*

Proof. Per Lemma 2.5.3, φ_k is strictly convex when $k > 0$, so its derivative in Lemma 2.4.3 is an increasing function over \mathbb{T} that is minimized at $t = -1$ and maximized at $t = 1$ (which is also true if $k = 0$). It is already proven by induction in Mathlib that $U_j(-1) = (-1)^j(j+1)$, which is apparent from Definition 7. Per Lemma 2.4.3, $\dot{\varphi}_k(-1) = \frac{1}{2} - \frac{1}{2} \sum_{m=1}^k 4^{-m} 2^m = \frac{1}{2} - \frac{1}{2} \sum_{m=1}^k 2^{-m} = 2^{-k-1}$ by Lemma D with $b = 2$. Lemma 2.5.1 then implies $\dot{\varphi}_k(1) = 1 - 2^{-k-1}$. \square

Lemma 2.5.6. *φ is a strictly increasing function over \mathbb{T} .*

Proof. Per Lemma 2.5.5, φ_k is a strictly increasing function over \mathbb{T} , and per Lemma 2.3.1 $\{\varphi_k\}$ converges uniformly to φ . The uniform limit of strictly increasing functions is a strictly increasing function, so φ is a strictly increasing function. \square

Lemma 2.5.7. *φ is a weak contraction over \mathbb{T} .*

Proof. Lemma 2.5.5 implies φ_k is Lipschitz continuous with a constant of $1 - 2^{-k-1}$, and per Lemma 2.3.1 $\{\varphi_k\}$ converges uniformly to φ . The uniform limit of Lipschitz continuous functions is Lipschitz continuous. Although the derivative of φ does not exist at $t = 1$, the supremum of the derivative of φ is 1, making it a weak contraction on \mathbb{T} . \square

Lemma 2.5.8. *If $t \in \mathbb{T}$, then $0 \leq \varphi_k(t) \leq 1$.*

Proof. Per Lemma 2.5.5, φ_k is a strictly increasing function over \mathbb{T} , which is minimized at $t = -1$ and maximized at $t = 1$. If $t = 1$, Lemma C implies $T_{2^m}(1) = 1$. Thus, Definition 2 specializes to $\varphi_k(1) = \frac{3}{7} + \frac{8^{-k}}{14} + \frac{1}{2} \sum_{m=0}^k 8^{-m} = \frac{3}{7} + \frac{8^{-k}}{14} + \frac{8-8^{-k}}{14} = 1$ by Lemma D with $b = 8$. Lemma F then implies $\varphi_k(-1) = 0$. \square

Hence, the unusual constant, $\frac{8^{-k}}{14}$, in Definition 2 anchors its range to be the same for all k .

Lemma 2.5.9. *If $t \in \mathbb{T}$, then $0 \leq \varphi(t) \leq 1$.*

Proof. Since the bounds on φ_k in Lemma 2.5.8 do not depend on k , they are preserved as $k \uparrow \infty$. \square

This chapter has demonstrated that our construction threads a needle. Lemma 2.3.2 states that φ is technically not continuously differentiable over \mathbb{T} , but only because it has essential singularities at both endpoints of \mathbb{T} . Thus, it is more useful to focus on the facts that φ is twice differentiable on $(-1, 1)$ per Lemma 2.4.9 and can be uniformly approximated by high-degree polynomials per Lemma 2.3.1. Thus, from a numerical perspective, φ is no different from the inverse trigonometric functions that are implemented in standard libraries, in the sense that they can be approximated to machine precision and the lack of a derivative at two or three points is a minor inconvenience.

Chapter 3

Injectivity

The goal of this chapter is to prove that Ψ is an injective function on \mathbb{I}^n for some $\lambda > 0$.

Definition 8. Let $\Psi(\mathbf{x}) \equiv \begin{bmatrix} \Psi_0(\mathbf{x} - 0) \\ \Psi_1(\mathbf{x} - \frac{1}{2n}) \\ \vdots \\ \Psi_q(\mathbf{x} - \frac{q}{2n}) \\ \vdots \\ \Psi_{2n}(\mathbf{x} - 1) \end{bmatrix}$ be a column vector of size $2n + 1$, where $\mathbf{x}^\top = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_p \\ \vdots \\ x_{n-1} \end{bmatrix}$

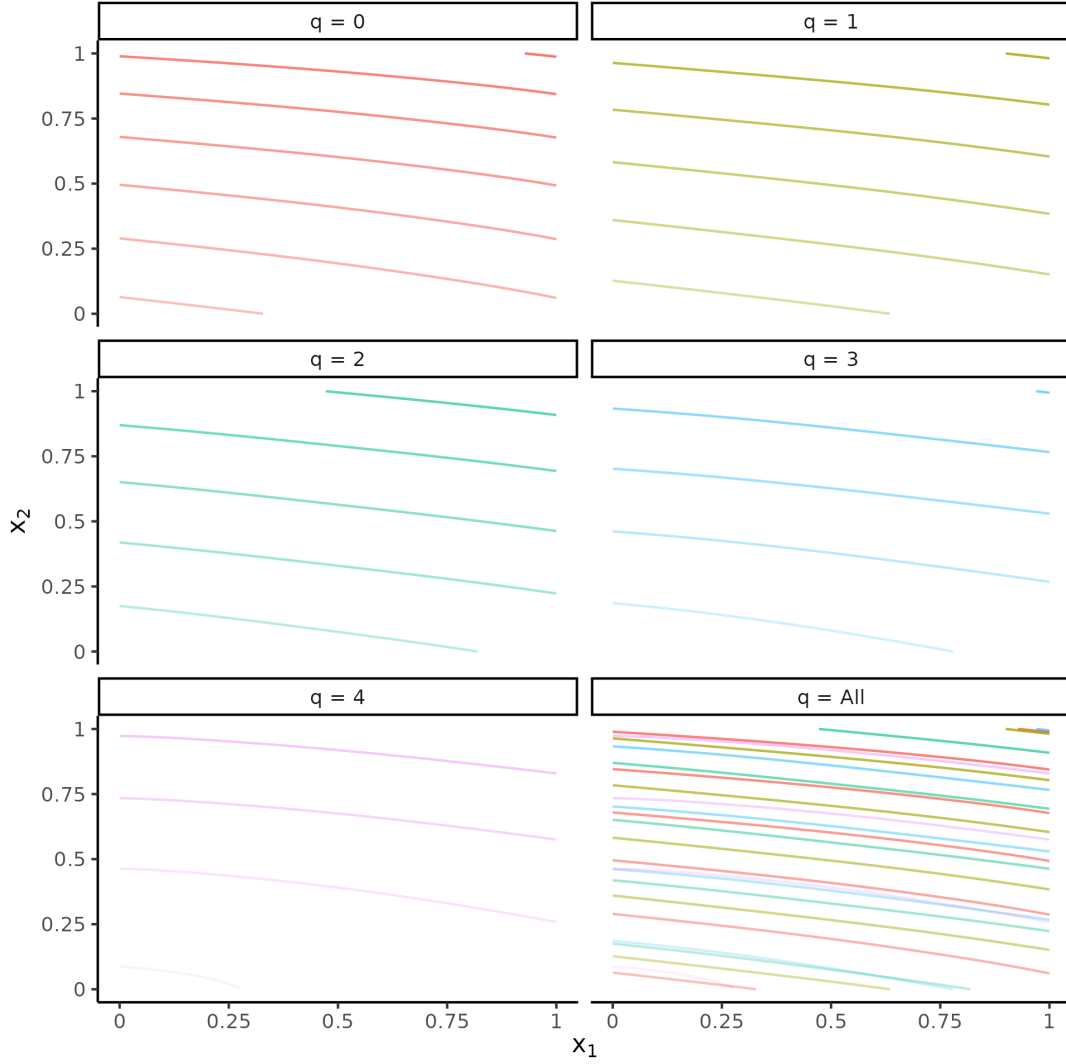
is a column vector of size n that holds the arguments to f , and $\Psi_q(\mathbf{x} - \frac{q}{2n}) \equiv \sum_{p=0}^{n-1} \frac{\lambda^p}{\Lambda} \varphi(x_p - \frac{q}{2n})$ in our construction from Chapter 2.

When $n = 2$, we can illustrate the issue as in Figure 3.1. For each of the five values of q , contour lines are plotted that depict pairs of values for x_1 and x_2 where $\Psi_q([x_1 \ x_2]) = \varphi(x - \frac{q}{2n})$ for various values of x . Sprecher's mechanism of shifting x_p by $-\frac{q}{2n}$ ensures that these contour lines are largely parallel to each other. Pairs of contour lines can and do intersect, but the KST cannot hold if all $2n + 1$ functions intersect at the same \mathbf{x} . There are an infinite number of level sets, each corresponding to a different value of x , so it is impossible to determine visually whether this condition is met. However, at the top right of the last plot, these contour lines are compressed into a small region, so the exact nature of the curvature there is important.

The contour lines are downward sloping and cross the diagonal, which justifies thinking of a level set as all possible vectors \mathbf{x} that render $\Psi_q(\mathbf{x})$ equal to $\varphi(x - \frac{q}{2n})$, which is the value that Ψ_q would take if all its arguments were equal to x . Sprecher

What this says about the superposition representation of an arbitrary continuous function $f(x_1, x_2)$ is that it is determined at five points on the diagonal through its representation ...The first thing to notice is the relation between the level sets of superpositions and target functions. An arbitrary function $f(x_1, x_2)$ may have infinitely many values on one or or more level sets ...But on these the functions $[\Psi_q]$ can only compute one value of f . [78 - 79]

Figure 3.1: Some level sets of $\Psi_q([x_1 \ x_2])$ when $n = 2$ and $\lambda = 2\pi - 1$



Each contour line is a level set, albeit at a different value of x , where $\Psi_q([x_1 \ x_2]) = \varphi(x - \frac{q}{2n})$.

3.1 Linear Algebra

Definition 9 (Weak Jacobian matrix). Let $\mathbf{J}(\mathbf{x})$ be a matrix with $2n + 1$ rows and n columns such that each element is $J_{q,p} \equiv \lambda_p \dot{\varphi}_\infty(x_p - \frac{q}{2n}) = \frac{\lambda^p}{2\Lambda} \sum_{m=0}^{\infty} 4^{-m} U_{-1+2m}(x_p - \frac{q}{2n})$ per Lemma 2.4.5.

Remark. $\mathbf{J}(\mathbf{x})$ in Definition 9 can be evaluated at a point in \mathbb{I}^n even when $\frac{\partial \Psi_q}{\partial x_p}$ is not defined, which occurs either when both $x_p = 1$ and $q = 0$ or when both $x_p = 0$ and $q = 2n$. Otherwise, $\mathbf{J}(\mathbf{x})$ is a traditional Jacobian matrix because the weak derivative coincides with the derivative.

Lemma 3.1.1. $\mathbf{J}(\mathbf{x})$ has full column rank for all $\mathbf{x} \in \mathbb{I}^n$.

Proof. The p -th column of $\mathbf{J}(\mathbf{x})$ is a nonlinear function only of x_p , so it is impossible to express any column as an exact linear function of the other columns. \square

Proposition 3.1.2. Ψ is an injective function on \mathbb{I}^n for any $\lambda > 0$

Proof. Lemma 3.1.1 suffices to prove that Ψ is locally injective throughout the interior of \mathbb{I}^n . Thus, we only need to rule out some global and boundary concerns. Per Lemma 2.5.7, φ is Lipschitz continuous and so is each $\psi_{p,q}$. Lipschitz continuous functions are absolutely continuous and so the Fundamental Theorem of Calculus applies. Suppose on the contrary that $\Psi(\mathbf{x}) = \Psi(\mathbf{x}')$ but $\mathbf{x} \neq \mathbf{x}'$. There is a smooth path between \mathbf{x} and \mathbf{x}' . Integrating the directional derivatives along that path should yield a value zero, but cannot be zero since $\mathbf{J}(\mathbf{x})$ has full column rank. \square

Remark. I am not entirely sure how the details of the previous “proof” would work, so below is another proof that is very consistent with the KST literature.

3.2 Trigonometric Representation

Lemma 3.2.1. If $t \in \mathbb{T}$, then $T_j(t) = \cos(j \cos^{-1} t)$.

Proof. This lemma has been known for decades and is already proven in Mathlib. Let $t = \cos \theta$. If $j = 0$, then $T_0(t) = 1 = \cos(0\theta)$. If $j = 1$, then $T_1(t) = t = \cos(1\theta)$. It remains to show for $j > 1$ that the three-term recursion from Definition 1 is satisfied: $T_j(t) = 2 \cos \theta \cos((j-1)\theta) - \cos((j-2)\theta)$. The first term can be rewritten as $2 \cos \theta \cos((j-1)\theta) = \cos(\theta - (j-1)\theta) + \cos(\theta + (j-1)\theta) = \cos((j-2)\theta) + \cos(j\theta)$, so $T_j(t) = \cos(j\theta)$. \square

Remark. Under Definition 3 and Lemma 3.2.1, $\frac{3}{7} + \frac{1}{2} \sum_{m=0}^{\infty} (\frac{1}{8})^m \cos(2^m \theta)$ resembles the function,

$w(y) = \sum_{m=0}^{\infty} \alpha^m \cos(\beta^m \pi y)$, Weierstrass proved to be nowhere differentiable when $\beta \geq 7$ also satisfies $\alpha\beta > 1 + \frac{3}{2}\pi$ for $0 < \alpha < 1$. But the terms in our φ have less extreme oscillation, smaller amplitudes, and another composition, so φ merely has essential singularities at the endpoints of \mathbb{T} .

Lemma 3.2.2 (Chebyshev-Lobatto nodes). The $1 + 2^k$ extrema of T_{2^k} are called (Chebyshev-Lobatto) nodes and occur at $\tau = -\cos \frac{j\pi}{2^k}$ with $0 \leq j \leq 2^k$ and $k \in \mathbb{N}$.

Proof. If $\tau = -\cos \frac{j\pi}{2^k}$, then $T_{2^k}(\tau) = \cos(2^k \cos^{-1}(\cos \frac{-j\pi}{2^k})) = \cos(-j\pi)$. If j is even, $T_{2^k}(\tau) = 1$ and if j is odd $T_{2^k}(\tau) = -1$. \square

The function approximation literature also defines nodes when the index of the Chebyshev polynomial is not a power of two, but we only consider the case where $\tau = -\cos \frac{j\pi}{2^k}$. It is essential to recognize that if $\tau = -\cos \frac{j\pi}{2^k}$, then it is also true that $\tau = -\cos \frac{(2j)\pi}{2^{k+1}}$. Thus, a node at iteration k is also a node at iteration $k+1$ but with an index that is double its previous index.

Definition 10 (γ -nodes). Let $\gamma = -\cos \frac{(3j+1)\pi}{2^{k-1} \times 3}$ for some $0 \leq j \leq 2^{k-1}$ and $k > 0$.

The orange points in Figure 2.1 are the values of $\varphi(\mp \frac{1}{2})$, whose argument $-\cos \frac{\pi}{3} = -\frac{1}{2}$ and $-\cos \frac{2\pi}{3} = \frac{1}{2}$ is a γ -node when $k = 1$; see the second remark following Lemma 2.3.1. The red points in Figure 2.1 are the values of φ at Chebyshev-Lobatto nodes when $k = 4$, which crucially can be calculated *exactly* in a finite number of arithmetic operations from φ_4 or better.

Lemma 3.2.3 (Zero approximation error). *If $\tau = -\cos \frac{j\pi}{2^k}$, then $\varphi_k(\tau) = \varphi(\tau)$.*

Proof. If $\tau = -\cos \frac{j\pi}{2^k}$, then Lemma 3.2.2 implies $T_{2^k}(\tau) = \mp 1$, and Lemma C implies $T_{2^{k+1}}(\tau) = 1$ is a fixed point. By Lemma D with $b = 8$, $\varphi(\tau) = \frac{3}{7} + \frac{1}{2} \sum_{m=0}^k 8^{-m} T_{2^m}(\tau) + \frac{1}{2} \sum_{m=k+1}^{\infty} 8^{-m} = \frac{3}{7} + \frac{1}{2} \sum_{m=0}^k 8^{-m} T_{2^m}(\tau) + \frac{8^{-k}}{14} = \varphi_k(\tau)$. Thus, the first-order correction, $\frac{8^{-k}}{14}$, is exact when τ is a node. \square

3.3 Number Theory

Lemma 3.3.1. *If $\tau = -\cos \frac{j\pi}{2^k}$ for some $0 < j < 2^k$, then τ is a root of U_{-1+2^k} .*

Proof. Standard using Lemma 2.4.1; extrema of polynomials are roots of their derivatives. \square

Definition 11 (Cyclotomic polynomial of order 2^m). Let $m > 0$ be an integer. The 2^m -th cyclotomic polynomial is the irreducible polynomial, $\Phi_{2^m}(r) \equiv r^{2^{m-1}} - 1$.

Lemma 3.3.2. *If $k > 0$, then $U_{-1+2^k}(r) = \prod_{m=1}^k \Phi_{2^m}(r)$.*

Proof. It can be done by induction. \square

Remark. U_{-1+2^k} factors into a product of irreducible polynomials, each of which only has complex roots, but their real parts are the roots of U_{-1+2^k} .

Definition 12 (Cyclotomic field). A cyclotomic field is an extension of the field of rational numbers, \mathbb{Q} , that is formed by adjoining some complex root, ζ , of a cyclotomic polynomial.

Definition 13 (Maximal real cyclotomic extension of \mathbb{Q} for power-of-2 roots of unity). Let $\mathbb{F}_\tau \equiv \mathbb{R} \cap \bigcup_{k=1}^{\infty} \mathbb{Q}(\zeta_{2^k})$, where ζ_{2^k} is some root of Φ_{2^k} .

Lemma 3.3.3. *If $\lambda \notin \mathbb{F}_\tau$, then Ψ_q is an injective function on $(\mathbb{F}_\tau \cap \mathbb{D})^n$.*

Proof. Let each $x_p - \frac{q}{2n}$ and $x'_p - \frac{q}{2n}$ be a member of $\mathbb{F}_\tau \cap \mathbb{I}$. Suppose $\Psi_q(x_0, x_1, \dots, x_{n-1}) = \Psi_q(x'_1, x'_2, \dots, x'_n)$ and rearrange this condition as

$$\frac{\varphi(x_0 - \frac{q}{2n}) - \varphi(x'_0 - \frac{q}{2n})}{\Lambda} = \frac{\sum_{p=1}^{n-1} \lambda^p [\varphi(x'_p - \frac{q}{2n}) - \varphi(x_p - \frac{q}{2n})]}{\Lambda}.$$

Suppose, on the contrary, that Ψ_q is not injective, which is to say that the above equation has a non-trivial solution. Then, the Mean Value Theorem states for a function like φ that is continuous on a closed interval (per Lemma 2.3.4) and differentiable on the open interval that excludes its endpoints (per Lemma 2.4.5), there exists a t_0 strictly between $x_0 - \frac{q}{2n}$ and $x'_0 - \frac{q}{2n}$ such that $\varphi(x_0 - \frac{q}{2n}) - \varphi(x'_0 - \frac{q}{2n}) = [x_0 - \frac{q}{2n} - x'_0 + \frac{q}{2n}] \dot{\varphi}(t_0)$. Moreover, t_0 is unique in this case because φ is a strictly increasing function on \mathbb{T} per Lemma 2.5.6. After canceling the Λ from the above equation, substituting with the Mean Value Theorem, and dividing both sides by $\dot{\varphi}(t_0)$, we obtain

$$x_0 - x'_0 = \frac{1}{\dot{\varphi}(t_0)} \sum_{p=1}^{n-1} \lambda^p \left[\varphi\left(x'_p - \frac{q}{2n}\right) - \varphi\left(x_p - \frac{q}{2n}\right) \right].$$

The left-hand side is a member of \mathbb{F}_λ , which cannot be equal to the right-hand side when $\lambda \notin \mathbb{F}_\tau$ because the latter is a member of $\mathbb{F}_\tau(\lambda, \dot{\varphi}(t_0)) / \mathbb{F}_\tau$. Thus, there is only a trivial solution to the original equation where $x_p = x'_p$ for all p , which characterizes an injective function. \square

Remark. Suppose $x_p - \frac{q}{2n} = -\cos \frac{j\pi}{2^k}$ and $x'_p - \frac{q}{2n} = -\cos \frac{j'\pi}{2^k}$, in which case the Mean Value Theorem implies there is a (in this case unique) t_p strictly between $x_p - \frac{q}{2n}$ and $x'_p - \frac{q}{2n}$ such that $[x_p - x'_p] \dot{\varphi}(t_p) = \varphi_k(x_p - \frac{q}{2n}) - \varphi_k(x'_p - \frac{q}{2n})$, where we have used Lemma 3.2.3 to substitute φ_k for φ . Although $\dot{\varphi}(t_p) \in \mathbb{F}_\tau$ because $x_p - x'_p \in \mathbb{F}_\tau$ and $\varphi_k(x_p - \frac{q}{2n}) - \varphi_k(x'_p - \frac{q}{2n}) \in \mathbb{F}_\tau$, it seems that $t_p \notin \mathbb{F}_\tau$ when $k > 1$. Thus, $\dot{\varphi}_k$ can somehow map some foreign numbers in \mathbb{T} into \mathbb{F}_τ .

Proposition 3.3.4. *For some $\lambda > 0$, Ψ is an injective function on \mathbb{I}^n .*

Proof. The fact that Ψ is injective on a dense set, $(\mathbb{F}_\tau \cap \mathbb{I})^n$, and is both convex (since it is a weighed sum of convex inner functions) and continuous allows us to do a proof by contradiction. \square

Chapter 4

Outer Functions

This chapter currently contains some disconnected fragments of thoughts.

Kahane proves that if there were fewer than $2n + 1$ rows in the output, then almost no mapping from \mathbb{I}^n would be injective and thus the KST would not hold, which is a result that is usually credited to Sternfeld a few years later. In addition, Whitney previously proved that a smooth mapping from a space of dimension n to a space of dimension $2n + 1$ can be further embedded into a space of dimension $2n$, which provides some intuition for Vitushkin and Khenkin's result that the KST cannot hold if all the inner functions were continuously differentiable.

Kaufman proves a mercifully short proof that the KST cannot hold if the genuine Jacobian matrix is of full rank, but this proof presumes on the equicontinuity of the inner functions, which does not hold in our case because $\psi_{p,q} \notin C^1(\mathbb{I})$ if $q = 0$ or $q = 2n$ (per Proposition 2.3.3), while $\psi_{p,q} \in C^2(\mathbb{I})$ otherwise (per Proposition 2.4.10). So, the question becomes is equicontinuity actually necessary for Kaufman's theorem?

The number of distinct outer functions can also be reduced to one by shifting the argument.

Definition 14. $\chi_q(\Psi_q(x_0, x_1, \dots, x_{n-1})) \equiv \chi(\Psi_q(x_0, x_1, \dots, x_{n-1}) + q)$.

Kahane's geometric interpretation of the KST goes like this. Let Γ_p denote an "increasing curve" in \mathbb{R}^{2n+1} for the equation $X_q(x) = \lambda_p \varphi(x - \frac{q}{2n})$. Let $E \equiv \lambda_1 \Gamma + \lambda_2 \Gamma + \dots + \lambda_n \Gamma$, so $E \subset \mathbb{I}^{2n+1}$. Then,

1. The mapping $\Gamma_1 \times \Gamma_2 \times \dots \times \Gamma_n$ is injective so E is a distorted cube
2. E is an interpolation set in the sense that every continuous function on can be written in the form $\chi_1(X_1) + \chi_2(X_2) + \dots + \chi_{2n}(X_{2n})$.

Definition 15 (Helson set). "[A] Helson set in a locally compact abelian group G is a closed set E such that the algebra $A(E)$ of restrictions to E of functions in $A(G)$ (Fourier transforms of summable functions on the dual group) coincides with $C(E)$, the algebra of all continuous functions on E ." (translated from page 91 of Kahane 1978)

Lemma 4.0.1. "If E contains the algebraic sum of two infinite sets, E is not a Helson set."

Proof. See page 91 of Kahane 1978; it has to do with violating a mesh condition. □

Lemma 4.0.2. *If γ coincides on any subinterval of \mathbb{I} with a polynomial of degree less than n , then $\{\gamma(\Psi_1), \gamma(\Psi_2), \dots, \gamma(\Psi_{2n})\}$ is not a Helson set.*

Proof. Also has do to with the mesh condition. □

Proposition 4.0.3. *If γ is an increasing homeomorphism of \mathbb{I} that does not coincide with a polynomial of degree less than n on any subinterval of \mathbb{I} , then $\gamma(E)$ is quasi-surely a Helson set. Moreover, the condition on γ is necessary and sufficient, and $\chi_q \in A^+(\mathbb{I})$ where $A^+(\mathbb{I})$ is a class of functions $\chi(z) = \sum_{j=0}^{\infty} \hat{a}_j e^{2\pi i j z}$.*

Proof. “The proof relies on the following proposition, which is very similar to one given in [Kahane’s 1975 article] p. 233 and can be proven in the same way.”

Let $B(\mathbb{I})$ be a Banach space contained in $C(\mathbb{I})$ and satisfying hypothesis (H): there exists $c > 0$ such that, for any $\delta > 0$, there exist finite sets $D_p \subset I$ (for $p = 1, 2, \dots, n$) with at least one point in every subinterval of length δ , such that the mapping $D_1 \times D_2 \times \dots \times D_n \rightarrow D_1 + D_2 + \dots + D_n$ is injective, and for any function u of modulus 1 on the set $D_1 + D_2 + \dots + D_n$, there exists $\chi \in B(\mathbb{I})$, with $\|\chi\|_{B(\mathbb{I})} \leq c$ and $\|\chi\|_{C(\mathbb{I})} = 1$, where $\chi = u$ on $D_1 + D_2 + \dots + D_n$. Then one can choose $\chi \in B(\mathbb{I})$.

[I]t follows [from Kronecker’s theorem] that any function of modulus 1 on \mathbb{I} can be written as $\sum_{j=1}^{\infty} a_j e^{i j z}$, with $\sum_{j=1}^{\infty} |a_j| \leq 2$. By applying the proposition, we can therefore take $\chi \in (A \circ \gamma)(\mathbb{I})$ as soon as homeomorphism γ satisfies the condition:

For every $\delta > 0$, there exist finite sets $D_p \subset \mathbb{I}$ (for $p = 1, 2, \dots, n$) at least one point in every subinterval of \mathbb{I} of length δ , such that the mapping $D_1 \times D_2 \times \dots \times D_n \rightarrow D_1 + D_2 + \dots + D_n$ is injective, and $\gamma(D_1 + D_2 + \dots + D_n)$ is a rationally independent set. □

$$\text{Then } \chi_q(\gamma(\Psi_q(x_0, x_1, \dots, x_{n-1}))) = \sum_{j=0}^{\infty} \hat{a}_j e^{2\pi i j (\gamma(\Psi_q(x_0, x_1, \dots, x_{n-1}))) + q} = \sum_{j=0}^{\infty} \hat{a}_j e^{2\pi i j (\gamma(\Psi_q(x_0, x_1, \dots, x_{n-1})))}$$

because it is periodic. In practice, we would have to estimate a (truncated) sequence of coefficients to calibrate to a particular target function.

Chapter 5

Open Problems

There are several that I have not written up yet.

Bibliography

- [1] Jonas Actor. *Computation for the Kolmogorov Superposition Theorem*. PhD thesis, 2018.
- [2] Jonas Actor and Matthew G Knepley. An algorithm for computing lipschitz inner functions in kolmogorov’s superposition theorem. *arXiv preprint arXiv:1712.08286*, 2017.
- [3] Buma L Fridman. An improvement in the smoothness of the functions in an kolmogorov’s theorem on superpositions. In *Doklady Akademii Nauk*, volume 177, pages 1019–1022. Russian Academy of Sciences, 1967.
- [4] T Hedberg. The kolmogorov superposition theorem, appendix ii to hs shapiro, topics in approximation theory. *Lecture notes in Math*, 187:267–275, 1971.
- [5] Jean-Pierre Kahane. Sur le theoreme de superposition de kolmogorov. *Journal of Approximation Theory*, 13(3):229–234, 1975.
- [6] Jean-Pierre Kahane. Baire’s category theorem and trigonometric series. *Journal D’Analyse Mathématique*, 80(1):143–182, 2000.
- [7] JP Kahane. Sur le treizième problème de hilbert, le théorème de superposition de kolmogorov et les sommes algébriques d’arcs croissants, in harmonic analysis iraklion 1978 proceedings, 1980.
- [8] S. Ya Khavinson. *Best approximation by linear superpositions (approximate nomography)*, volume 159. American Mathematical Soc., 1997.
- [9] Andrei Nikolaevich Kolmogorov. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. In *Doklady Akademii Nauk*, volume 114, pages 953–956. Russian Academy of Sciences, 1957.
- [10] Hadrien Montanelli and Haizhao Yang. Error bounds for deep relu networks using the kolmogorov–arnold superposition theorem. *Neural Networks*, 2020.
- [11] David A Sprecher. An improvement in the superposition theorem of kolmogorov. *Journal of Mathematical Analysis and Applications*, 38(1):208–213, 1972.
- [12] Lloyd N Trefethen. *Approximation theory and approximation practice, extended edition*. SIAM, 2019.
- [13] Anatoliy Georgievich Vitushkin and GM Khenkin. Linear superpositions of functions. *Russian Mathematical Surveys*, 22(1):77, 1967.