A. Mathematical Remarks

A.1. Infinite Sums

Throughout this paper we consider expressions of the following form:

$$\Phi(X) = \Sigma_{x \in X} \phi(x) \tag{5}$$

Where X is an arbitrary set. The meaning of this expression is clear when X is finite, but when X is infinite, we must be precise about what we mean.

A.1.1. Countable Sums

We usually denote countable sums as e.g. $\Sigma_{i=1}^{\infty} x_i$. Note that there is an ordering of the x_i here, whereas there is no ordering in our expression (5). The reason that we consider sums is for their permutation invariance in the finite case, but note that in the infinite case, permutation invariance of sums does not necessarily hold! For instance, the alternating harmonic series $\Sigma_{i=1}^{\infty} \frac{(-1)^i}{i}$ can be made to converge to any real number simply by reordering the terms of the sum. For expressions like (5) to make sense, we must require that the sums in question are indeed permutation invariant. This property is known as *absolute convergence*, and it is equivalent to the property that the sum of absolute values of the series converges. So for (5) to make sense, we will require everywhere that $\Sigma_{x \in X} |\phi(x)|$ is convergent. For any X where this is not the case, we will set $\Phi(X) = \infty$.

A.1.2. Uncountable Sums

It is well known that a sum over an uncountable set of elements only converges if all but countably many elements are 0. Allowing sums over uncountable sets is therefore of little interest, since it essentially reduces to the countable case.

A.2. Continuity of Functions on Sets

We are interested in functions on subsets of $\mathbb{R}$, i.e. elements of $2^{\mathbb{R}}$, and the notion of continuity on $2^{\mathbb{R}}$ is not straightforward. As a convenient shorthand, we discuss "continuous" functions f on $2^{\mathbb{R}}$, but what we mean by this is that the function f_M induced by f on $\mathbb{R}^M$ by $f_N(x_1, ..., x_M) = f(\{x_1, ..., x_M\})$ is continuous for every $M \in \mathbb{N}$.

A.3. Remark on Theorem 2.8

The proof for Theorem 2.8 from Zaheer et al. (2017) can be extended to dealing with multi sets, i.e. sets with repeated elements. To that end, we replace the mapping to natural numbers $c(X) : \mathbb{R}^M \to \mathbb{N}$ with a mapping to prime numbers $p(X) : \mathbb{R}^M \to \mathbb{P}$. We then choose $\phi(x_m) = -\log p(x_m)$. Therefore,

$$\Phi(X) = \sum_{m=1}^{M} \phi(x_m) = \log \prod_{m=1}^{M} \frac{1}{p(x_m)} \tag{6}$$

which takes a unique value for each distinct X therefore extending the validity of the proof to multi-sets. However, unlike the original series, this choice of ϕ diverges with infinite set size.

In fact, it is straightforward to show that there is no function ϕ for which Φ provides a unique mapping for arbitrary multi-sets while at same time guaranteeing convergence for infinitely large sets. Assume a function ϕ and an arbitrary point x such that $\phi(x) = a \neq 0$. Then, the multiset comprising infinitely many identical members x would give:

$$\Phi(X) = \sum_{i=1}^{\infty} \phi(x_m) = \sum_{i=1}^{\infty} a = \pm\infty \tag{7}$$

B. Proofs of Theorems

B.1. Theorem 3.1

Theorem 3.1. *There exist functions $f : 2^{\mathbb{Q}} \to \mathbb{R}$ such that, whenever (ρ, ϕ) is a sum-decomposition of f via $\mathbb{R}$, ϕ is discontinuous at every point $q \in \mathbb{Q}$.*

Proof. Consider $f(X) = \sup(X)$, the least upper bound of X. Write $\Phi(X) = \Sigma_{x \in X} \phi(x)$. So we have:

$$\sup(X) = \rho(\Phi(X))$$

First note that $\phi(q) \neq 0$ for any $q \in \mathbb{Q}$. If we had $\phi(q) = 0$, then we would have, for every $X \subset \mathbb{Q}$:

$$\Phi(X) = \Phi(X) + \phi(q) = \Phi(X \cup \{q\})$$

But then, for instance, we would have:

$$q = \sup(\{q-1, q\}) = \sup(\{q-1\}) = q-1$$

This is a contradiction, so $\phi(q) \neq 0$.

Next, note that $\Phi(X)$ must be finite for every upper-bounded $X \subset \mathbb{Q}$ (since sup is undefined for unbounded X, we do not consider such sets, and may allow Φ to diverge). Even if we allowed the domain of ρ to be $\mathbb{R} \cup \{\infty\}$, suppose $\Phi(X) = \infty$ for some upper-bounded set X. Then:

$$\begin{aligned}\sup(X) &= \rho(\Phi(X)) \\ &= \rho(\infty) \\ &= \rho(\infty + \phi(\text{sup(X)} + 1)) \\ &= \rho(\Phi(X \cup \{sup(X) + 1\})) \\ &= \sup(X \cup \{sup(X) + 1\}) \\ &= sup(X) + 1\end{aligned}$$

This is a contradiction, so $\Phi(X) < \infty$ for any upper-bounded set X.

Now from the above it is immediate that, for any upper-bounded set X, only finitely many $x \in X$ can have $\phi(x) > \frac{1}{n}$. Otherwise we can find an infinite upper-bounded set $Y \subset X$ with $\phi(y) > \frac{1}{n}$ for every $y \in Y$, and $\Phi(Y) = \infty$.

Finally, let $q \in \mathbb{Q}$. We have already shown that $\phi(q) \neq 0$, and we will now construct a sequence q_n with:

1. $q_n \to q$
2. $\phi(q_n) \to 0$

If ϕ were continuous at q, we would have $\phi(q_n) \to \phi(q)$, so the above two points together will give us that ϕ is discontinuous at q.

So now, for each $n \in \mathbb{N}$, consider the set B_n of points which lie within $\frac{1}{n}$ of q. Since only finitely many points $p \in B_n$ have $\phi(p) > \frac{1}{n}$, and B_n is infinite, there must be a point $q_n \in B_n$ with $\phi(q_n) < \frac{1}{n}$. The sequence of such q_n clearly satisfies both points above, and so ϕ is discontinuous everywhere. □

B.2. Theorem 3.2

Theorem 3.2. *Let $f : \mathbb{R}^{\mathcal{F}} \to \mathbb{R}$. Then f is sum-decomposable via $\mathbb{R}$.*

Proof. Define $\Phi : \mathbb{R}^{\mathcal{F}} \to \mathbb{R}$ by $\Phi(X) = \Sigma_{x \in X}\phi(x)$. If we can demonstrate that there exists some ϕ such that Φ is injective, then we can simply choose $\rho = f \circ \Phi^{-1}$ and the result is proved.

Say that a set $X \subset \mathbb{R}$ is *finite-sum-distinct* if, for any finite subsets $A, B \subset X$, $\Sigma_{a \in A} a \neq \Sigma_{b \in B} b$. Now, if we can show that there is a finite-sum-distinct set D with the same cardinality as $\mathbb{R}$ (we denote $|\mathbb{R}|$ by $\mathfrak{c}$), then we can simply choose ϕ to be a bijection from $\mathbb{R}$ to D. Then, by finite-sum-distinctness, Φ will be injective, and the result is proved.

Now recall the statement of Zorn's Lemma: suppose $\mathcal{P}$ is a partially ordered set (or *poset*) in which every totally ordered subset has an upper bound. Then $\mathcal{P}$ has a maximal element.

The set of f.s.d. subsets of $\mathbb{R}$ (which we will denote $\mathcal{D}$) forms a poset ordered by inclusion. Supposing that $\mathcal{D}$ satisfies the conditions of Zorn's Lemma, it must have a maximal element, i.e. there is a f.s.d. set D_{max} such that any set E with $D_{\text{max}} \subsetneq E$ is not f.s.d. We claim that D_{max} has cardinality $\mathfrak{c}$.

To see this, let D be a f.s.d. set with infinite cardinality $\kappa < \mathfrak{c}$ (any maximal D clearly cannot be finite). We will show that $D \neq D_{\text{max}}$. Define the *forbidden elements* with respect to D to be those elements x of $\mathbb{R}$ such that $D \cup \{x\}$ is not f.s.d. We denote this set of forbidden elements F_D. Now note that, if D is maximal, then $D \cup F_D = \mathbb{R}$. In particular, this implies that $|F_D| = \mathfrak{c}$. But now consider the elements of F_D. By definition of F_D, we have that $x \in F_D$ if and only if $\exists c_1, ..., c_m, d_1, ..., d_n \in D$ such that $c_1 + ... + c_m + x = d_1 + ... + d_n$. So we can write x as a sum of finitely many elements of D, minus a sum of finitely many other elements of D. So there is a surjection from pairs of finite sets of D to elements of F_D. i.e.:

$$|F_D| \leq |D^{\mathcal{F}} \times D^{\mathcal{F}}|$$

But since D is infinite:

$$|D^{\mathcal{F}} \times D^{\mathcal{F}}| = |D| = \kappa < \mathfrak{c}$$

So $|F_D| < \mathfrak{c}$, and therefore $|D|$ is not maximal. This demonstrates that D_{max} must have cardinality $\mathfrak{c}$.

To complete the proof, it remains to show that $\mathcal{D}$ satisfies the conditions of Zorn's Lemma, i.e. that every totally ordered subset (or *chain*) $\mathcal{C}$ of $\mathcal{D}$ has an upper bound. So consider:

$$C_{\text{ub}} = \bigcup \mathcal{C} = \bigcup_{C \in \mathcal{C}} C$$

We claim that C_{ub} is an upper bound for $\mathcal{C}$. It is clear that $C \subset C_{\text{ub}}$ for every $C \in \mathcal{C}$, so it remains to be shown that $C_{\text{ub}} \in \mathcal{D}$, i.e. that C_{ub} is f.s.d.

We proceed by contradiction. Suppose that C_{ub} is not f.s.d. Then:

$$\exists c_1, ..., c_m, d_1, ..., d_n \in C_{\text{ub}} : \Sigma_i c_i = \Sigma_j d_j \tag{8}$$

But now by construction of C_{ub} there must be sets $C_1, ..., C_m, D_1, ..., D_m \in \mathcal{C}$ with $c_i \in C_i, d_j \in D_j$. Let $\mathcal{B} = \{C_i\}_{i=1}^m \cup \{D_j\}_{j=1}^n$. $\mathcal{B}$ is totally ordered by inclusion and all sets contained in it are f.s.d., since it is a subset of $\mathcal{C}$. Since $\mathcal{B}$ is finite it has a maximal element B_{max}. By maximality, we have $c_i, d_j \in B_{\text{max}}$ for all c_i, d_j. But then by (8), B_{max} is not f.s.d., which is a contradiction. So we have that C_{ub} is f.s.d.

In summary:

1. $\mathcal{D}$ satisfies the conditions of Zorn's Lemma.
2. Therefore there exists a maximal f.s.d. set, D_{max}.
3. We have shown that any such set must have cardinality $\mathfrak{c}$.
4. Given an f.s.d. set D_{max} with cardinality $\mathfrak{c}$, we can choose ϕ to be a bijection between $\mathbb{R}$ and D_{max}.
5. Given such a ϕ, we have that $\Phi(X) = \Sigma_{x \in X} \phi(x)$ is injective on $R^{\mathcal{F}}$.
6. Given injective Φ, choose $\rho = f \circ \Phi^{-1}$.
7. This choice gives us $f(X) = \rho(\Sigma_{x \in X} \phi(x))$ by construction.

This completes the proof. □

B.3. Theorem 3.3

Theorem 3.3. *If $\mathfrak{X}$ is uncountable, then there exist functions $f : 2^{\mathfrak{X}} \to \mathbb{R}$ which are not sum-decomposable. Note that this holds even if the sum-decomposition (ρ, ϕ) is allowed to be discontinuous.*

Proof. Consider $f(X) = \sup(X)$.

As discussed above, a sum over uncountably many elements can converge only if countably many elements are non-zero. But as in the proof of Theorem 3.1, $\phi(x) \neq 0$ for any x. So it is immediate that sum-decomposition is not possible for functions operating on uncountable subsets of $\mathfrak{X}$.

Even restricting to countable subsets is not enough. As in the proof of Theorem 3.1, we must have that for each $n \in \mathbb{N}$, $\phi(x) > \frac{1}{n}$ for only finitely many x. But then if this is the case, let $\mathfrak{X}_n$ be the set of all $x \in \mathfrak{X}$ with $\phi(x) > \frac{1}{n}$. Since $\phi(x) \neq 0$, we know that $\mathfrak{X} = \bigcup \mathfrak{X}_n$. But this is a countable union of finite sets, which is impossible because $\mathfrak{X}$ is uncountable.

□

B.4. Theorem 4.3

Theorem 4.3 (Fixed set size)**.** *Let $f : \mathbb{R}^M \to \mathbb{R}$ be continuous. Then f is permutation-invariant if and only if it is continuously sum-decomposable via $\mathbb{R}^M$.*

Proof. The reverse implication is clear. The proof relies on demonstrating that the function $\Phi \to \mathbb{R}^{M+1}$ defined as follows is a homeomorphism onto its image:

$$\Phi_q(X) = \sum_{m=1}^{M} \phi_q(x_m), \quad q = 0, \ldots, M$$
$$\phi_q(x) = x^q, \quad q = 0, \ldots, M$$

Now define $\widetilde{\Phi} \to \mathbb{R}^M$ by:

$$\widetilde{\Phi}_q(X) = \sum_{m=1}^{M} \widetilde{\phi}_q(x_m), \quad q = 1, \ldots, M$$
$$\widetilde{\phi}_q(x) = x^q, \quad q = 1, \ldots, M$$

Note that $\Phi_0(X) = M$ for all X, so $\text{Im}(\Phi) = \{M\} \times \text{Im}(\widetilde{\Phi})$. Since $\{M\}$ is a singleton, these two images are homeomorphic, with a homeomorphism given by:

$$\gamma : \text{Im}(\widetilde{\Phi}) \to \text{Im}(\Phi)$$
$$\gamma(x_1, \ldots, x_M) = (M, x_1, \ldots, x_M)$$

Now by definition, $\widetilde{\Phi} = \gamma^{-1} \circ \Phi$. Since this is a composition of homeomorphisms, $\widetilde{\Phi}$ is also a homeomorphism. Therefore $(f \circ \widetilde{\Phi}^{-1}, \widetilde{\phi})$ is a continuous sum-decomposition of f via $\mathbb{R}^M$. □

B.5. Theorem 4.4

Theorem 4.4 (Variable set size)**.** *Let $f : \mathbb{R}^{\leq M} \to \mathbb{R}$ be continuous. Then f is permutation-invariant if and only if it is continuously sum-decomposable via $\mathbb{R}^M$.*

Proof. We use the adapted sum-of-power mapping $\widetilde{\Phi}$ from above, denoted in this section by Φ.

$$\Phi_q(X) = \sum_{m=1}^{M} \phi_q(x_m), \quad q = 1, \ldots, M$$
$$\phi_q(x_m) = (x_m)^q, \quad q = 1, \ldots, M$$

which is shown above to be injective. Without loss of generality, let $\mathfrak{X} = [0, 1]$ as in Theorem 2.9.

We separate $\Phi_q(X)$ into two terms:

$$\Phi_q(X) = \sum_{m=1}^{M'} \phi_q(x_m) + \sum_{m=M'+1}^{M} \phi_q(x_m) \qquad (9)$$

For an input set X with $M' = M - P$ elements and $0 \leq M', P \leq M$, we say that the set contains M' "actual elements" as well as P "empty" elements which are not in fact part of the input set. Those P "empty elements" can be regarded as place fillers when the size of the input set is smaller than M, i.e. $M' < M$.

We map those P elements to a constant value $k \notin \mathfrak{X}$, preserving the injectiveness of $\Phi_q(X)$ for input sets X of arbitrary size M':

$$\Phi_q(X) = \sum_{m=1}^{M'} \phi_q(x_m) + \sum_{m=M'+1}^{M} \phi_q(k) \tag{10}$$

Equation (10) is no longer strictly speaking a sum-decomposition. This can be overcome by re-arranging it:

$$\begin{aligned} \Phi_q(X) &= \sum_{m=1}^{M'} \phi_q(x_m) + \sum_{m=M'+1}^{M} \phi_q(k) \\ &= \sum_{m=1}^{M'} \phi_q(x_m) + \sum_{m=1}^{M} \phi_q(k) - \sum_{m=1}^{M'} \phi_q(k) \\ &= \sum_{m=1}^{M'} [\phi_q(x_m) - \phi_q(k)] + \sum_{m=1}^{M} \phi_q(k) \end{aligned} \tag{11}$$

The last term in Equation (11) is a constant value which only depends on the choice of k and is independent of X and M'. Hence, we can replace $\phi_q(x)$ by $\widehat{\phi_q}(x) = \phi_q(x) - \phi_q(k)$. This leads to a new sum-of-power mapping $\widehat{\Phi}_q(X)$ with:

$$\begin{aligned} \widehat{\Phi}_q(X) &= \sum_{m=1}^{M'} \widehat{\phi}_q(x_m) \\ &= \Phi_q(X) - M \cdot \phi_q(k) \end{aligned} \tag{12}$$

$\widehat{\Phi}$ is injective since Φ is injective, $k \notin \mathfrak{X}$, and the last term in the above sum is constant. $\widehat{\Phi}$ is also in the form of a sum-decomposition.

For each $m < M$, we can follow the reasoning used in the rest of the proof of Theorem 2.9 to note that $\widehat{\Phi}$ is a homeomorphism when restricted to sets of size m – we denote these restricted functions by $\widehat{\Phi}_m$. Now each $\widehat{\Phi}_m^{-1}$ is a continuous function into $\mathbb{R}^m$. We can associate with each a continuous function $\widehat{\Phi}_{m,M}^{-1}$ which maps into $\mathbb{R}^M$, with the $M - m$ tailing dimensions filled with the value k.

Now the domains of the $\widehat{\Phi}_{m,M}^{-1}$ are compact and disjoint since $k \notin \mathfrak{X}$. We can therefore find a function $\widehat{\Phi}_C^{-1}$ which is continuous on $\mathbb{R}^N$ and agrees with each $\widehat{\Phi}_{m,M}^{-1}$ on its domain.

To complete the proof, let $\mathcal{Y}$ be a connected compact set with $k \in \mathcal{Y}, \mathfrak{X} \subset \mathcal{Y}$. Let $\widehat{f}$ be a function on subsets of $\mathcal{Y}$ of size exactly M satisfying:

$$\widehat{f}(X) = f(X); \quad X \subset \mathfrak{X}$$
$$\widehat{f}(X) = f(X \cap \mathfrak{X}); \quad X \subset \mathfrak{X} \cup \{k\}$$

We can choose $\widehat{f}$ to be continuous under the notion of continuity in Appendix A.2. Then $(\widehat{f} \circ \widehat{\Phi}_C^{-1}, \widehat{\phi})$ is a continuous sum-decomposition of f.

□

B.6. Max-Decomposition

Analogously to sum-decomposition, we define the notion of *max-decomposition*. A function f is max-decomposable if there are functions ρ and ϕ such that:

$$f(\mathbf{x}) = \rho\big(\max_i(\phi(x_i))\big).$$

where the max is taken over each dimension independently in the latent space. Our definitions of decomposability via Z and continuous decomposability also extend to the notion of max-decomposition.

We now state and prove a theorem which is closely related to Theorem 4.1, but which establishes limitations on max-decomposition, rather than sum-decomposition.

Theorem B.1. *Let $M > N \in \mathbb{N}$. Then there exist permutation invariant continuous functions $f : \mathbb{R}^M \to \mathbb{R}$ which are not max-decomposable via $\mathbb{R}^N$.*

Note that this theorem rules out any max-decomposition, whether continuous or discontinuous. We specifically demonstrate that summation is not max-decomposable – as with Theorem 4.1, this theorem applies to ordinary well-behaved functions.

Proof. Consider $f(\mathbf{x}) = \sum_{i=1}^{M} x_m$. Let $\phi : \mathbb{R} \to \mathbb{R}^N$, and let $\mathbf{x} \in \mathbb{R}^M$ such that $x_i \neq x_j$ when $i \neq j$.

For $n = 1, \ldots, N$, let $\mu(n) \in \{1, \ldots, M\}$ such that:

$$\max_i(\phi(x_i)_n) = \phi(x_{\mu(n)})_n$$

That is, $\phi(x_{\mu(q)})$ attains the maximal value in the q-th dimension of the latent space among all $\phi(x_i)$. Now since $N < M$, there is some $m \in \{1, \ldots, M\}$ such that $\mu(n) \neq m$ for any $n \in \{1, \ldots, N\}$. So now consider $\widetilde{\mathbf{x}}$ defined by:

$$\widetilde{x}_i = x_i; i \neq m \tag{13}$$
$$\widetilde{x}_m = x_{\mu(1)} \tag{14}$$

Then:

$$\max_i(\phi(x_i)) = \max_i(\phi(\widetilde{x}_i))$$

But since we chose $\mathbf{x}$ such that all x_i were distinct, we have $\sum_{i=1}^{M} x_i \neq \sum_{i=1}^{M} \widetilde{x}_i$ by the definition of $\widetilde{\mathbf{x}}$. This shows that ϕ cannot form part of a max-decomposition for f. But ϕ was arbitrary, so no max-decomposition exists.

□

C. A Continuous Function on $\mathbb{Q}$

This section defines and analyses the function Ψ shown in Figure 2, which is continuous on $\mathbb{Q}$ but not on $\mathbb{R}$. Ψ is defined as the pointwise limit of a sequence of functions Ψ_n, illustrated in Figure 4. We proceed as follows:

1. Define a sequence of functions $\widetilde{\Psi}_n$ on $[0, 1]$.
2. Show that the pointwise limit $\widetilde{\Psi}$ is continuous except at points of the form $k \cdot 2^{-m}$ for some integers k and m, i.e. except at the *dyadic rationals*.
3. Define the function Ψ on $[0, A]$ by $\Psi(x) = \widetilde{\Psi}(\frac{x}{A})$.
4. Note that Ψ is continuous except at points of the form $A \cdot k \cdot 2^{-m}$ for some integers k and m.
5. Choose A to be irrational, so that all points of discontinuity are also irrational, to obtain a function which is continuous on $\mathbb{Q}$. (In all figures, we have chosen $A = \log(4)$).

Informally, we set $\widetilde{\Psi_0}(x) = x$, and at iteration n, we split the unit interval into 2^n even subintervals. In every even-numbered subinterval, we reflect the function horizontally around the midpoint of the subinterval. We may write this formally as follows.

Let $x \in [0, 1], n \in \mathbb{N}$. Let:

$$a_n(x) = \frac{\lceil x \cdot 2^n \rceil + \frac{1}{2}}{2^n}$$

That is, $a_n(x)$ is the midpoint of the unique half-open interval containing x:

$$\Big(k \cdot 2^{-n}, (k+1) \cdot 2^{-n}\Big]; \quad k \in \mathbb{N}$$

Write $b_n(x)$ for the n-th digit in the binary expansion of x, and write $c_n(x)$ for the number of $b_m(x), m \leq n$ with $b_m(x) = 1$.

Importantly, $b_n(x)$ is ambiguous if x is a dyadic rational, since in this case x has both a terminating and a non-terminating expansion. For consistency with our choice of the upward-closed interval for the definition of $a_n(x)$, we choose the non-terminating expansion in this case.

Then:

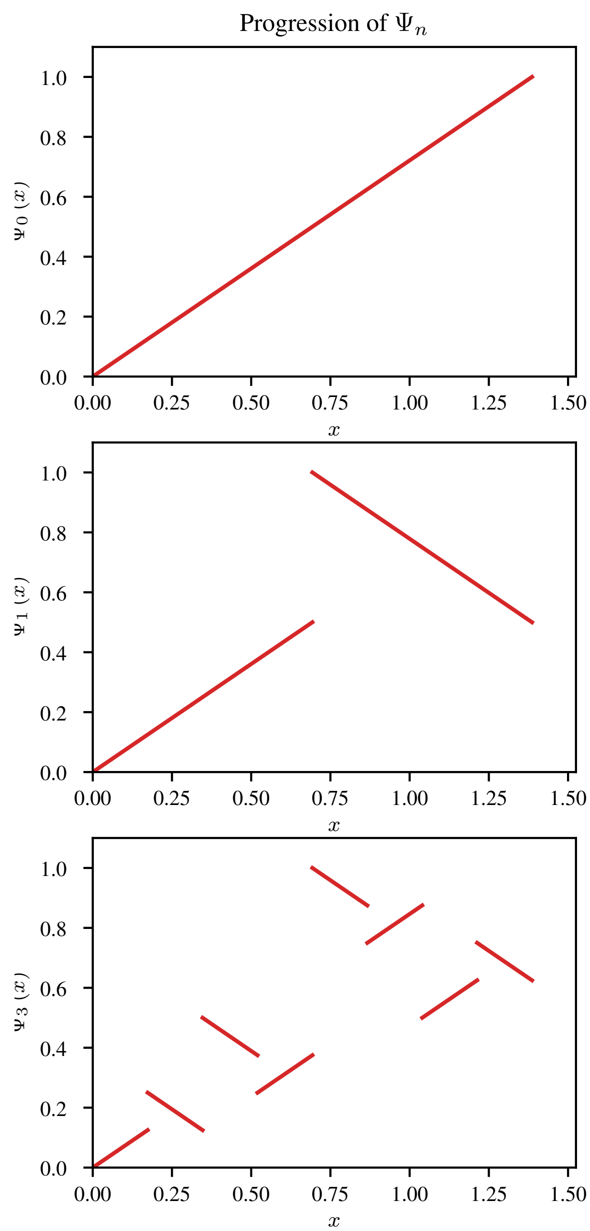

Figure 4: Several iterations of Ψ_n

$$\widetilde{\Psi}_n(x) = x + \sum_{i=1}^{n} (-1)^{c_i(x)} \cdot b_i(x) \cdot 2(x - a_i(x))$$
$$\widetilde{\Psi}(x) = x + \sum_{i=1}^{\infty} (-1)^{c_i(x)} \cdot b_i(x) \cdot 2(x - a_i(x))$$

First, it is clear that the series for $\widetilde{\Psi}(x)$ converges absolutely at every x, since $|2(x - a_i(x))| \leq 2^{-i}$. So this function is well defined. Also note that:

$$|\widetilde{\Psi}_n(x) - \widetilde{\Psi}(x)| \leq 2^{-n} \tag{15}$$

Note further that $\widetilde{\Psi}_n$ is continuous except at points of the form $k \cdot 2^{-n}$, since a_m, b_m and c_m are continuous at these points for all $m \leq n$.

Now consider a point x_* which is not a dyadic rational. We wish to show that $\widetilde{\Psi}$ is continuous at x_*. So let $\epsilon > 0$. Choose n so that $2^{-n} < \frac{\epsilon}{3}$. Since x_* is not a dyadic rational, $\widetilde{\Psi}_n$ is continuous at x_*, i.e. there is some $\delta > 0$ such that $|x_* - x| < \delta \implies |\widetilde{\Psi}_n(x_*) - \widetilde{\Psi}_n(x)| < \frac{\epsilon}{3}$. But now, by Equation (15):

$$|\widetilde{\Psi}(x_*) - \widetilde{\Psi}_n(x_*)| < 2^{-n} < \frac{\epsilon}{3}$$
$$|\widetilde{\Psi}(x) - \widetilde{\Psi}_n(x)| < 2^{-n} < \frac{\epsilon}{3}$$

And so, whenever $|x_* - x| < \delta$, we have:

$$\begin{aligned} |\widetilde{\Psi}(x_*) - \widetilde{\Psi}(x)| &\leq |\widetilde{\Psi}(x_*) - \widetilde{\Psi}_n(x_*)| \\ &\quad + |\widetilde{\Psi}_n(x_*) - \widetilde{\Psi}_n(x)| \\ &\quad + |\widetilde{\Psi}_n(x) - \widetilde{\Psi}(x)| \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} \\ |\widetilde{\Psi}(x_*) - \widetilde{\Psi}(x)| &< \epsilon \end{aligned}$$

Thus, $\widetilde{\Psi}$ is continuous at x_*.

As noted above, we may now set $\Psi(x) = \widetilde{\Psi}(\frac{x}{A})$ to obtain a function which is continuous at all rational points.

D. Implementation Details for Illustrative Example

The network setup follows Equation (1) with 3 fully connected layers before the summation (acting on each input independently) and 2 fully connected layers after. Each fully connected layer has 1000 hidden units and is followed by a ReLU non-linearity. However, the third hidden layer, which creates the latent space in which the summation is executed, has a variable dimension N. N is varied for each experiment in order to examine the influence of the latent dimension on the performance.

Training was conducted using the ADAM optimizer (Kingma & Ba, 2015) with an initial learning rate of 0.001 and an exponential decay after each batch of 0.99. Training was ended after convergence at 500 batches with a batch size of 32. Samples were continuously drawn from the respective distributions. Therefore, there is no notion of training vs. test data or epoch sizes. The results r^t, measured as RMSE, were smoothed using exponential smoothing with $\alpha = 0.95$:

$$r^t_{\text{smooth}} = (1 - \alpha) \cdot r^t + \alpha \cdot r^{t-1} \tag{16}$$

The last, smoothed RMSE is extracted from each experiment, averaged over 500 different runs with different seeds and plotted in Figure 3(a). The confidence intervals are calculated assuming a Gaussian distribution. The critical points are extracted by taking the smallest latent dimension which produces an RMSE of less than 10% above the global minimum for this set size.

Out of distribution samples: We tested the performance of a trained model (500 inputs, 100 latent dimensions) on out-of-distribution samples to see to what extent the model exploits the statistical properties of the training examples. Below is a list with examples:

- Input: $[1.0, 1.0, \ldots, 1.0]$, output: 0.933, true label: 1.0
- Input: 200 times 1.0 and 300 times 0.0, output: 0.413, true label: 0.0
- Input: $[0.002, 0.004, 0.006, ..., 1.0]$, output: 0.5006, true label: 0.5000

The distributions the samples were drawn from during test time were the uniform distribution, a Gaussian and a Gamma distribution. Each sample consisting of 500 values was randomly drawn from one of these distributions. Hence, the first two are very unlikely samples from the provided distributions. The poor performance of the model on these two examples can therefore be taken as an indication that the model does utilize information about the underlying distributions when estimating the median. The third sample is much closer to a realistic sample from, e.g., the uniform distribution (in our case between 0.0 and 1.0), which makes it unsurprising that the model performs much better on this task. It is worth noting that the notion of 'likely' examples of uniform distributions is of course an intuitive one. Given that no value is repeated, every specific set of numbers is of course equally likely as long as all numbers lie within the interval of the uniform distribution.