PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY, SERIES B
Volume 10, Pages 296–310 (September 22, 2023)
https://doi.org/10.1090/bproc/186

PIECEWISE LINEAR FUNCTIONS REPRESENTABLE WITH INFINITE WIDTH SHALLOW RELU NEURAL NETWORKS

SARAH MCCARTY

(Communicated by Yuan Xu)

Abstract. This paper analyzes representations of continuous piecewise linear functions with infinite width, finite cost shallow neural networks using the rectified linear unit (ReLU) as an activation function. Through its integral representation, a shallow neural network can be identified by the corresponding signed, finite measure on an appropriate parameter space. We map these measures on the parameter space to measures on the projective n-sphere cross $\mathbb{R}$, allowing points in the parameter space to be bijectively mapped to hyperplanes in the domain of the function. We prove a conjecture of Ongie et al. [*A Function Space View of Bounded Norm Infinite Width ReLU Nets: The Multivariate Case*, arXiv, 2019] that every continuous piecewise linear function expressible with this kind of infinite width neural network is expressible as a finite width shallow ReLU neural network.

1. Introduction

We consider shallow neural networks which use rectified linear unit (ReLU) as the activation function. It is well known ReLU has universal approximation properties on compact domains, and in practice has advantages over sigmoidal activation functions [6, 16]. Finite width shallow neural networks with $n+1$-dimensional input take the form

$$f(\boldsymbol{x}) = c_0 + \sum_{i=1}^{k} c_i \sigma\left(\boldsymbol{a}_i \cdot \boldsymbol{x} - b_i\right) \tag{1}$$

where $\boldsymbol{a}_i \in \mathbb{S}^n$ (the unit sphere in $\mathbb{R}^{n+1}$) and $b_i, c_i \in \mathbb{R}$ for all i.

Generalizing to infinite width neural networks transforms the sum to an integral and the weights c_i to a signed measure μ on $\mathbb{S}^n \times \mathbb{R}$ where

$$f(\boldsymbol{x}) = \int_{\mathbb{S}^n \times \mathbb{R}} \sigma\left(\boldsymbol{a} \cdot \boldsymbol{x} - b\right) \mathrm{d}\mu(\boldsymbol{a}, b) + c_0. \tag{2}$$

Some authors choose instead to have the measure and integral over all of $\mathbb{R}^{n+1} \times \mathbb{R}$. An important class of functions are those representable with an infinite width neural network with finite representation cost, which corresponds with $|\mu|\left(\mathbb{S}^n \times \mathbb{R}\right) < \infty$ [3]. Similar classes of functions are studied in [7] as Barron spaces and in [2] as $\mathcal{F}_1$.

To ensure the integral in Equation 2 is well defined, we can require μ has a finite first moment where $\int_{\mathbb{S}^n \times \mathbb{R}} |b| \,\mathrm{d}|\mu|(\boldsymbol{a}, b) < \infty$. Alternatively, Ongie et al. in [15]

Received by the editors October 6, 2022, and, in revised form, July 17, 2023.
2020 *Mathematics Subject Classification.* Primary 68T07, 42C40, 41A30.
The author was supported by the National Science Foundation (No. 1934884).

writes the integral in the form

$$\int_{\mathbb{S}^n \times \mathbb{R}} \sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b) - \sigma(-b)\, \mathrm{d}\mu(\boldsymbol{a}, b) + c_0, \tag{3}$$

so the integral is well-defined whenever μ is a finite measure. Since they differ only by a constant, the class of functions representable with a finite measure in the form of Equation 2 is a subclass of the class of functions representable with a finite measure in the form of Equation 3 [15]. Therefore, we choose to consider integral representations in the form of Equation 3.

Since $\sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b)$ is a ridge function, integral representations are also naturally studied as the dual ridgelet transform on distributions [4, 14, 18]. Functions with these representations are then often analyzed with the Radon transform [14, 15, 18]. Savarese in [17] characterized which one-dimensional functions are representable with infinite width, finite cost shallow ReLU neural networks.

Many finitely piecewise linear functions cannot be represented with a finite width ReLU shallow network, including all non-trivial compactly supported piecewise linear functions on $\mathbb{R}^n$, $n \geq 2$ [15]. Lower and upper bounds on number of layers needed to represent continuous piecewise linear functions with finite width, deep neural networks have been established [1, 11]. However, the class of infinite width ReLU networks is certainly more expressive than the class of finite width networks in general, such as being able to express some non-piecewise linear functions [17]. It is not obvious if the class of infinite width shallow ReLU neural networks can express a finitely piecewise linear function that the class of finite width shallow ReLU networks cannot. By decomposing measures, E and Wojtowytsch in [8] established the set of points of non-differentiability of a function in a Barron space must be a subset of a countable union of affine subspaces. However, proper subsets are possible. In [15], Ongie et al. proved many compactly supported piecewise linear functions are not representable with finite cost, infinite width shallow ReLU neural networks. This led to Conjecture 1.

Conjecture 1 (Ongie et al., [15])**.** *A continuous piecewise linear function f has finite representation cost if and only if it is exactly representable by a finite width shallow neural network.*

A finite representation cost corresponds with the existence of a finite measure μ such that f admits a representation in the form of Equation 3. Our main result is to prove the conjecture, which for precision we formulate here as a theorem.

Theorem. *Let $f : \mathbb{R}^{n+1} \to \mathbb{R}$ be a continuous finitely piecewise linear function. If there exist a finite, signed Borel measure μ on $\mathbb{S}^n \times \mathbb{R}$ and $c_0 \in \mathbb{R}$ such that $f(\boldsymbol{x}) = \int_{\mathbb{S}^n \times \mathbb{R}} \sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b) - \sigma(-b)\, \mathrm{d}\mu(\boldsymbol{a}, b) + c_0$, then f is representable as a finite-width network as in Equation 1.*

This result will be a corollary of Theorem 1, which will be stated after establishing notation.

1.1. Notation.

For $m \in \mathbb{N}$, let $[m] \coloneqq \{1, \ldots, m\}$.

The rectified linear unit (ReLU) function from $\mathbb{R}$ to $\mathbb{R}$ is denoted $\sigma(t)$ and defined as $\sigma(t) \coloneqq \max\{0, t\}$.

The pushforward measure of measure μ induced by a mapping φ is denoted $\mu \circ \varphi^{-1}$.

Regular lower case Latin letter variables generally represent real numbers: $x, y, z, t \in \mathbb{R}$. Once n is fixed, bold lower case Latin letter variables indicate elements of $\mathbb{R}^{n+1}$: $\boldsymbol{a}, \boldsymbol{x} \in \mathbb{R}^{n+1}$. Bold lower case Greek letter variables indicate elements of $\mathbb{R}^n$: $\boldsymbol{\zeta}, \boldsymbol{\xi} \in \mathbb{R}^n$.

Let $\mathfrak{h}_{\boldsymbol{a},b} \coloneqq \{\boldsymbol{x} \in \mathbb{R}^{n+1} \mid \boldsymbol{a} \cdot \boldsymbol{x} = b\}$.

For $m \in \mathbb{N}$, $m-1$ dimensional affine subspaces in $\mathbb{R}^m$ are called hyperplanes. The m-sphere in $\mathbb{R}^{m+1}$ is denoted $\mathbb{S}^m$. The real-projective space is denoted $\mathbb{RP}^m$.

Let $\boldsymbol{e}_{m+1} \coloneqq (0, \dots, 0, 1) \in \mathbb{S}^m$.

Let $\mathcal{S}^0 \coloneqq \{1\} \subseteq \mathbb{S}^0$. For $m \geq 1$, let $\mathcal{S}^m$ be defined as

$$\mathcal{S}^m \coloneqq \{\boldsymbol{x} \in \mathbb{S}^m \mid \boldsymbol{e}_{m+1} \cdot \boldsymbol{x} > 0\} \cup \{(x_1, \dots, x_m, 0) \mid (x_1 \dots, x_m) \in \mathcal{S}^{m-1}\} \subseteq \mathbb{S}^m. \tag{4}$$

Let $-\mathcal{S}^m$ denote the pointwise negation of all the points in $\mathcal{S}^m$. By simple induction, exactly one of $\boldsymbol{x}, -\boldsymbol{x} \in \mathcal{S}^m$ for all $\boldsymbol{x} \in \mathbb{S}^m$. Therefore, $\mathbb{S}^m = \mathcal{S}^m \sqcup (-\mathcal{S}^m)$.

Let $D_{\boldsymbol{d}^+} f(\boldsymbol{x})$ denote the one-sided directional derivative of f in the positive direction of $\boldsymbol{d}$ for $\boldsymbol{d} \in \mathbb{S}^n$.

For any metric space W, $\mathcal{B}(W)$ denotes the set of Borel sets and $\mathcal{M}(W)$ denotes the set of Borel, finite, signed measures on W.

1.2. **Overview.** The key to proving the conjecture is the following theorem. Recall, a representation of $\mathbb{RP}^n$, $\mathcal{S}^n$, is precisely defined in Equation 4.

Theorem 1. *Suppose $\mu \in \mathcal{M}(\mathcal{S}^n \times \mathbb{R})$ is such that*

(1) *μ is atomless*
(2) *$f(\boldsymbol{x}) = \int_{\mathcal{S}^n \times \mathbb{R}} \sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b) - \sigma(-b) \, \mathrm{d}\mu(\boldsymbol{a}, b)$ is a continuous countably piecewise linear function from $\mathbb{R}^{n+1}$ to $\mathbb{R}$.*

Then, μ is the zero measure.

Informally, Theorem 1 states, with $\mathcal{S}^n \times \mathbb{R}$ as a parameter space, atomless measures cannot induce the sudden change in first-order derivatives that occur at the boundaries of affine pieces. Since point masses have easy-to-characterize effects on affineness, this implies the conjecture with $\mathcal{S}^n \times \mathbb{R}$ as the parameter space. Lemmas 4 and 5 show expanding the parameter space only introduces new affine terms to the representable functions. The conjecture will be established in Corollary 3.

The conjecture only concerns finitely piecewise linear functions and its proof only relies on Theorem 1 applied to finitely piecewise linear functions. However, the proof of Theorem 1 naturally extends to countably piecewise linear functions and provides insight into the role of point masses for this broader class of functions, particularly in Corollary 2.

Additionally, the conjecture implies non-trivial compactly supported piecewise linear functions in dimensions higher than two cannot be represented in a shallow ReLU network (Corollary 4). In contrast, many such functions are representable in finite-width two-layer networks, such as $f(x_1, x_2) = \sigma(1 - \sigma(2x_1) - \sigma(-x_1 + 2x_2) - \sigma(-x_1 - 2x_2))$.

2. Preliminaries

We start by formally defining countably piecewise linear.

Definition 1. A *convex polyhedron* C is a subset of $\mathbb{R}^n$ such that $C = \bigcap_{H \in \mathcal{H}} H$ where $\mathcal{H}$ is a finite set of closed half-spaces. A *defining supporting hyperplane* of C

with respect to $\mathcal{H}$ is a hyperplane $\mathfrak{h}$ that is the boundary of a half-space in $\mathcal{H}$ such that $C \cap \mathfrak{h} \neq \varnothing$.

Remark 1. For each polyhedron, there are many acceptable choices of $\mathcal{H}$ of varying cardinalities. We will assume there is a fixed choice and refer to the finite set of defining supporting hyperplanes of C. The boundary of C is always a closed subset of the union of defining supporting hyperplanes.

Definition 2. A *continuous countably (finitely) piecewise linear function* is a continuous function such that there is a countable (finite) collection of convex polyhedra that cover the domain where the function is affine when restricted to each polyhedron.

Remark 2. The requirement of continuity in the definition does not impose any limitations on the results. Every function with a representation of the form $f(\boldsymbol{x}) = \int_{\mathbb{S}^n \times \mathbb{R}} \sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b) - \sigma(-b)\, \mathrm{d}\mu(\boldsymbol{a}, b) + c_0$ is continuous.

In our proof, we use the fact first-order directional derivatives are constant on affine polyhedra. To simplify calculations, we will be particularly interested in directional derivatives in the direction $\boldsymbol{e}_{n+1} \coloneqq (0, \ldots, 0, 1)$.

Lemma 1. *Suppose* $f(\boldsymbol{x}) = \int_{\mathcal{S}^n \times \mathbb{R}} \sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b) - \sigma(-b)\, \mathrm{d}\mu(\boldsymbol{a}, b)$ *where* $\mu \in \mathcal{M}(\mathcal{S}^n \times \mathbb{R})$. *Then,*

$$D_{\boldsymbol{e}_{n+1}^{+}} f(\boldsymbol{x}) = \int_{\{(\boldsymbol{a},b) \in \mathcal{S}^n \times \mathbb{R} \,\mid\, \boldsymbol{a} \cdot \boldsymbol{x} \geq b\}} \boldsymbol{a} \cdot \boldsymbol{e}_{n+1}\, \mathrm{d}\mu(\boldsymbol{a}, b).$$

Proof. First,

$$\begin{aligned} &\lim_{h \to 0^+} \frac{f(\boldsymbol{x} + h\boldsymbol{e}_{n+1}) - f(\boldsymbol{x})}{h} \\ &= \lim_{h \to 0^+} \int_{\mathcal{S}^n \times \mathbb{R}} \frac{\sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b + \boldsymbol{a} \cdot (h\boldsymbol{e}_{n+1})) - \sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b)}{h}\, \mathrm{d}\mu(\boldsymbol{a}, b). \end{aligned}$$

Whenever $\boldsymbol{a} \cdot \boldsymbol{x} < b$, for sufficiently small h,

$$\sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b + \boldsymbol{a} \cdot (h\boldsymbol{e}_{n+1})) = \sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b) = 0.$$

Thus, when $\boldsymbol{a} \cdot \boldsymbol{x} < b$,

$$\lim_{h \to 0^+} \frac{\sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b + \boldsymbol{a} \cdot (h\boldsymbol{e}_{n+1})) - \sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b)}{h} = 0.$$

By definition of $\mathcal{S}^n$, $\boldsymbol{a} \cdot (h\boldsymbol{e}_{n+1}) \geq 0$ when $h \geq 0$ for all $\boldsymbol{a} \in \mathcal{S}^n$. Hence, if $\boldsymbol{a} \cdot \boldsymbol{x} - b \geq 0$, then $\boldsymbol{a} \cdot \boldsymbol{x} - b + \boldsymbol{a} \cdot (h\boldsymbol{e}_{n+1}) \geq 0$ for all $\boldsymbol{a} \in \mathcal{S}^n$ and $h \geq 0$. It follows whenever $\boldsymbol{a} \cdot \boldsymbol{x} \geq b$,

$$\begin{aligned} &\lim_{h \to 0^+} \frac{\sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b + \boldsymbol{a} \cdot (h\boldsymbol{e}_{n+1})) - \sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b)}{h} \\ &= \lim_{h \to 0^+} \frac{\boldsymbol{a} \cdot \boldsymbol{x} - b + \boldsymbol{a} \cdot (h\boldsymbol{e}_{n+1}) - (\boldsymbol{a} \cdot \boldsymbol{x} - b)}{h} = \lim_{h \to 0^+} \frac{\boldsymbol{a} \cdot (h\boldsymbol{e}_{n+1})}{h} = \boldsymbol{a} \cdot \boldsymbol{e}_{n+1}. \end{aligned}$$

Further, as $\|\boldsymbol{a}\| = \|\boldsymbol{e}_{n+1}\| = 1$, for all $\boldsymbol{a}, \boldsymbol{x}, b$ and all $h \geq 0$,

$$\left| \frac{\sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b + \boldsymbol{a} \cdot (h\boldsymbol{e}_{n+1})) - \sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b)}{h} \right| \leq \frac{|\boldsymbol{a} \cdot (h\boldsymbol{e}_{n+1})|}{h} \leq 1. \tag{5}$$

Since $|\mu|(\mathcal{S}^n \times \mathbb{R}) < \infty$, by Equation 5, the Dominated Convergence Theorem applies. Therefore,

$$\begin{aligned}
&\lim_{h\to 0^+} \frac{f(\boldsymbol{x} + (h\boldsymbol{e}_{n+1})) - f(\boldsymbol{x})}{h} \\
&= \int_{\mathcal{S}^n\times\mathbb{R}} \lim_{h\to 0^+} \frac{\sigma\left(\boldsymbol{a}\cdot\boldsymbol{x} - b + \boldsymbol{a}\cdot(h\boldsymbol{e}_{n+1})\right) - \sigma\left(\boldsymbol{a}\cdot\boldsymbol{x} - b\right)}{h} \,\mathrm{d}\mu\left(\boldsymbol{a}, b\right) \\
&= \int_{\{(\boldsymbol{a},b)\in\mathcal{S}^n\times\mathbb{R} \mid \boldsymbol{a}\cdot\boldsymbol{x}\geq b\}} \lim_{h\to 0^+} \frac{\boldsymbol{a}\cdot(h\boldsymbol{e}_{n+1})}{h} \,\mathrm{d}\mu\left(\boldsymbol{a}, b\right) \\
&= \int_{\{(\boldsymbol{a},b)\in\mathcal{S}^n\times\mathbb{R} \mid \boldsymbol{a}\cdot\boldsymbol{x}\geq b\}} \boldsymbol{a}\cdot\boldsymbol{e}_{n+1} \,\mathrm{d}\mu\left(\boldsymbol{a}, b\right). \qquad \square
\end{aligned}$$

The proof of Theorem 1 will proceed by induction on the dimension of the domain, therefore we prove the one-dimensional case first. While the one-dimensional case is much simpler than the general case, the technique of using first-order derivatives to conclude a measure is zero will be repeated.

Lemma 2. *Suppose $\mu \in \mathcal{M}(\mathcal{S}^0 \times \mathbb{R})$ is such that*

(1) *μ is atomless*
(2) *$f(x) = \int_{\mathcal{S}^0\times\mathbb{R}} \sigma\left(ax - b\right) - \sigma(-b) \,\mathrm{d}\mu\left(a, b\right)$ is a continuous countably piecewise linear function.*

Then, μ is the zero measure.

Proof. Since $|\mathcal{S}^0| = |\{1\}| = 1$, μ is uniquely determined by its marginal measure on $\mathbb{R}$, $\mu_{\mathbb{R}}$. As f is countably piecewise linear, there are countably many intervals $\{[q_i, r_i]\}_{i\in\mathbb{N}}$ such that f is affine when restricted to each interval and $\mathbb{R} \setminus \left(\bigcup_{i\in\mathbb{N}}(q_i, r_i)\right)$ is countable. Further, μ is atomless, so $\mu_{\mathbb{R}}$ is determined by its values on closed intervals that are subsets of intervals in $\{(q_i, r_i)\}_{i\in\mathbb{N}}$ [12]. Thus, it suffices to show $\mu_{\mathbb{R}}([x_1, x_2]) = 0$ whenever f is affine on (q, r) and $[x_1, x_2] \subseteq (q, r)$.

Suppose f is affine on (q, r) and $[x_1, x_2] \subseteq (q, r)$. By Lemma 1,

$$\begin{aligned}
f'(x_2) &= \int_{\{(a,b)\in\mathcal{S}^0\times\mathbb{R} \mid x_2\geq b\}} a\cdot 1 \,\mathrm{d}\mu\left(a, b\right) = \int_{\{(a,b)\in\mathcal{S}^0\times\mathbb{R} \mid x_2\geq b\}} 1\cdot 1 \,\mathrm{d}\mu\left(a, b\right) \\
&= \int_{\{(a,b)\in\mathcal{S}^0\times\mathbb{R} \mid x_2\geq b\}} 1 \,\mathrm{d}\mu\left(a, b\right) = \mu\left(\left\{(a, b) \in \mathcal{S}^0 \times \mathbb{R} \mid x_2 \geq b\right\}\right).
\end{aligned}$$

Similarly,

$$f'(x_1) = \mu\left(\left\{(a, b) \in \mathcal{S}^0 \times \mathbb{R} \mid x_1 \geq b\right\}\right).$$

Therefore, as f is affine in-between x_1 and x_2,

$$0 = f'(x_2) - f'(x_1) = \mu\left(\{1\} \times \left(x_1, x_2\right]\right) = \mu\left(\{1\} \times \left[x_1, x_2\right]\right) = \mu_{\mathbb{R}}([x_1, x_2]).$$

$\square$

3. Constructing a dense set of directions

Any set of countably many points has zero weight with respect to an atomless measure. In the one-dimensional case, this allows us to disregard points in $\mathcal{S}^0 \times \mathbb{R}$ associated with boundaries when determining μ is the zero measure. However, in higher dimensions there are more than countably many points associated with boundaries. Nonetheless, a carefully picked subset of points associated with boundaries will have zero weight with respect to μ and will be large enough to ultimately conclude μ is the zero measure. The first step to constructing this set is finding a large set of non-co-hyperplanar points.

Proposition 1. *For every $n \in \mathbb{N}$, there exists a set $S \subseteq \mathbb{R}^n$ such that*

(1) *For every open ball $B \subseteq \mathbb{R}^n$, $S \cap B$ is uncountable*
(2) *For every hyperplane $P \subseteq \mathbb{R}^n$, $|S \cap P| \leq n$.*

Proof. By [19], there is a set $I \subseteq \mathbb{R}$ algebraically independent over $\mathbb{Q}$ such that $|I| = |\mathbb{R}|$.

There is a bijective function $\phi : [n] \times \mathbb{N} \times \mathbb{R} \to I$.

Let $\{B_m\}_{m \in \mathbb{N}}$ be an enumeration of open balls in $\mathbb{R}^n$ centered at rational coordinates with rational radius.

Note, $0 \notin I$. For every $m \in \mathbb{N}, r \in \mathbb{R}$, there exist $q_{1,m,r}, \ldots, q_{n,m,r} \in \mathbb{Q} \setminus \{0\}$ such that

$$(q_{1,m,r}\phi(1, m, r), \ldots, q_{n,m,r}\phi(n, m, r)) \in B_m.$$

The set $\{q_{\ell,m,r}\phi(\ell, m, r) \mid \ell \in [n], m \in \mathbb{N}, r \in \mathbb{R}\}$ is also algebraically independent and each element has a unique representation of the form $q_{\ell,m,r}\phi(\ell, m, r)$. Define

$$S := \{(q_{1,m,r}\phi(1, m, r), \ldots, q_{n,m,r}\phi(n, m, r)) \mid m \in \mathbb{N}, r \in \mathbb{R}\}.$$

Consider an open ball $B \subseteq \mathbb{R}^n$. Since $\mathbb{Q}$ is dense, there is m_0 such that $B_{m_0} \subseteq B$. Further, $\{(q_{1,m_0,r}\phi(1, m_0, r), \ldots, q_{n,m_0,r}\phi(n, m_0, r)) \mid r \in \mathbb{R}\} \subseteq B_{m_0} \subseteq B$. It follows $S \cap B$ is uncountable.

By way of contradiction, suppose there exist distinct $\left(z_0^1, \ldots, z_0^n\right), \ldots, \left(z_n^1, \ldots, z_n^n\right) \in S \cap P$ for some hyperplane P. It follows any n vectors between these points are linearly dependent, so

$$\det \begin{bmatrix} z_0^1 - z_1^1 & \ldots & z_0^n - z_1^n \\ \vdots & \ddots & \vdots \\ z_0^1 - z_n^1 & \ldots & z_0^n - z_n^n \end{bmatrix} = 0. \tag{6}$$

The determinant is a polynomial over $\mathbb{Q}$ in terms of z_i^j. Since a unique z_i^j, $i \geq 1$, is an addend in each entry, the determinant cannot be the trivial polynomial. This contradicts the z_i^j being algebraically independent.

It follows for all hyperplanes P, $|S \cap P| \leq n$. ☐

Corollary 1. *Suppose $S \subseteq \mathbb{R}^n$ is as in Proposition 1. Let $\phi : S \to \mathbb{R}$ and $S' := \{(\boldsymbol{\zeta}, \phi(\boldsymbol{\zeta})) \mid \boldsymbol{\zeta} \in S\} \subseteq \mathbb{R}^{n+1}$. For every $n-1$ dimensional affine subspace $P \subseteq \mathbb{R}^{n+1}$, $|S' \cap P| \leq n$.*

Proof. Let $\phi : S \to \mathbb{R}$. By way of contradiction, let P be a $n-1$ dimensional affine subspace and suppose distinct points $(z_0^1, \ldots, z_0^{n+1}), \ldots, (z_n^1, \ldots, z_n^{n+1}) \in S' \cap P$. Then,

$$\operatorname{rank} \begin{bmatrix} z_0^1 - z_1^1 & \ldots & z_0^n - z_1^n \\ \vdots & \ddots & \vdots \\ z_0^1 - z_n^1 & \ldots & z_0^n - z_n^n \end{bmatrix} \leq \operatorname{rank} \begin{bmatrix} z_0^1 - z_1^1 & \ldots & z_0^{n+1} - z_1^{n+1} \\ \vdots & \ddots & \vdots \\ z_0^1 - z_n^1 & \ldots & z_0^{n+1} - z_n^{n+1} \end{bmatrix} \leq n - 1.$$

Therefore, as in Equation 6 of Proposition 1,

$$\det \begin{bmatrix} z_0^1 - z_1^1 & \ldots & z_0^n - z_1^n \\ \vdots & \ddots & \vdots \\ z_0^1 - z_n^1 & \ldots & z_0^n - z_n^n \end{bmatrix} = 0,$$

a contradiction. ☐

4. Proofs of main results

Lemma 3. *Let W be a metric space and $\mu \in \mathcal{M}(W)$. Consider a collection of Borel sets $\mathcal{P} \subseteq \mathcal{B}(W)$ such that there exists a $c \in \mathbb{N}$ where $|\mu|\left(\bigcap_{i\in[c]} P_i\right) = 0$ for all distinct $P_1, \dots, P_c \in \mathcal{P}$. Then, there are only countably many $P \in \mathcal{P}$ such that $|\mu|(P) > 0$.*

Proof. Every uncountable family of sets of positive measure has an infinite subfamily with positive intersection [10]. The lemma follows from the contrapositive. ☐

The sets in the following definition are the intersection of certain half-spaces in $\mathbb{R}^{n+1}$ and will be used in the proof of Theorem 1.

Definition 3. Suppose $\boldsymbol{\zeta}_0 \in \mathbb{R}^n, y_1, y_2 \in \mathbb{R} \cup \{\pm\infty\}$ with $y_1 \leq y_2$. Define

$$L_{\boldsymbol{\zeta}_0}(y_1, y_2) \coloneqq \{(\boldsymbol{\xi}, v) \in \mathbb{R}^n \times \mathbb{R} \mid v - \boldsymbol{\xi} \cdot \boldsymbol{\zeta}_0 \in (y_1, y_2]\}$$

and

$$\overline{L}_{\boldsymbol{\zeta}_0}(y_1) \coloneqq \{(\boldsymbol{\xi}, v) \in \mathbb{R}^n \times \mathbb{R} \mid v - \boldsymbol{\xi} \cdot \boldsymbol{\zeta}_0 = y_1\}.$$

We now outline the argument of Theorem 1 before the proof. Recall, Theorem 1 will be proved by induction on the dimension of the domain. The inductive step of Theorem 1 will show μ is zero everywhere on the interior of the parameter space. Then, f is constant as x_{n+1} changes and is characterized by a function of lower-dimension. Hence, the inductive hypothesis applies. The proof of the inductive step is divided into five parts.

(1) **Definitions and maps between measure spaces.** There is a natural bijection between the parameter space and hyperplanes in the domain. It links μ to where it induces changes in first-order derivatives. There is a related bijection φ between the interior of the parameter space and $\mathbb{R}^{n+1}$.
(2) **Refinement of S.** We generate large sets of points in the domain from S which lie on the boundaries of the affine pieces of f. For most of these points, the measure of the set associated with non-affineness at the point is zero. Since S is large, we can remove any problematic points and refine S to S'.
(3) **Vanishing integrals over line segments.** In Parts (3) and (4), we consider an integral with respect to $\mu \circ \varphi^{-1}$ of a function closely related to the first-order directional derivatives. In Part (3), we integrate over sets associated with non-affineness within certain line segments in the domain. These sets take the form $L_{\boldsymbol{\zeta}_0}(y_1, y_2)$, motivating Definition 3. When this line segment is entirely contained in a polyhedron on which f is affine, the integral vanishes.
(4) **Vanishing integrals over half-spaces.** We now consider the integral over sets associated with nonaffineness within certain rays in the domain. Through the maps in Part (1), these correspond with half-spaces in $\mathbb{R}^{n+1}$. The ray is broken into line segments entirely contained within polyhedra on which f is affine and points associated with S' on boundaries. Combining the results of the previous two parts shows these integrals vanish.

(5) **Conclusion with Cramer–Wold and Radon–Nikodym.** Since Part (4) concerns half-spaces, the Cramer–Wold theorem applies. Along with the Radon–Nikodym theorem, we can show $\mu \circ \varphi^{-1}$ on $\mathbb{R}^{n+1}$ is zero. Thus, μ is zero on the interior of the parameter space.

Proof of Theorem 1. First, Lemma 2 proves the theorem for the case $n = 0$. For induction, assume the theorem holds for $n - 1$.

Definitions and maps between measure spaces. Suppose $\mu \in \mathcal{M}\left(\mathcal{S}^n \times \mathbb{R}\right)$ satisfies all the hypotheses.

Let $\mathcal{C}$ be a countable collection of convex polyhedra that cover the domain of f such that f is affine on each.

Define the following sets of hyperplanes in $\mathbb{R}^{n+1}$

$$\mathcal{H}_0 \coloneqq \{\mathfrak{h}_{\boldsymbol{a},b} \mid \boldsymbol{a} \in \mathcal{S}^n, b \in \mathbb{R}\} \quad \text{and} \quad \mathcal{H}_1 \coloneqq \{\mathfrak{h}_{\boldsymbol{a},b} \mid \boldsymbol{a} \in \mathcal{S}^n, \boldsymbol{a} \cdot \boldsymbol{e}_{n+1} \neq 0, b \in \mathbb{R}\}.$$

Define the map $\gamma : \mathcal{S}^n \times \mathbb{R} \to \mathcal{H}_0$ as $\gamma\left(\boldsymbol{a}, b\right) = \mathfrak{h}_{\boldsymbol{a},b}$. By construction of $\mathcal{S}^n$, this is bijective.

Define the map $\psi : \mathcal{H}_1 \to \mathbb{R}^n \times \mathbb{R}$ such that for $\boldsymbol{a} = (a_1, \ldots, a_{n+1})$,

$$\psi\left(\mathfrak{h}_{\boldsymbol{a},b}\right) = \left(\frac{a_1}{a_{n+1}}, \ldots, \frac{a_n}{a_{n+1}}, \frac{b}{a_{n+1}}\right). \tag{7}$$

By definition of $\mathcal{H}_1$, it is routine to verify ψ is well-defined and bijective.

The image of ψ is $\mathbb{R}^n \times \mathbb{R}$, however, elements in the image of ψ should *not* be thought of as being in the domain of f. Therefore, identify generic elements in the image of ψ with $(\boldsymbol{\xi}, v) \in \mathbb{R}^n \times \mathbb{R}$ and call the space $\Xi \times V$ where $\Xi = \mathbb{R}^n$, $V = \mathbb{R}$.

Define $\varphi : \gamma^{-1}[\mathcal{H}_1] \to \Xi \times V$ as $\varphi \coloneqq \psi \circ \gamma$. Then, $\mu \circ \varphi^{-1}$ is a measure on $\Xi \times V$. Since φ is bijective, $\mu \circ \varphi^{-1}$ is atomless.

For fixed $\boldsymbol{\zeta}_0 \in \mathbb{R}^n$, $y_0 \in \mathbb{R}$,

$$\psi\left[\{\mathfrak{h}_{\boldsymbol{a},b} \in \mathcal{H}_1 \mid \boldsymbol{a} \cdot (\boldsymbol{\zeta}_0, y_0) = b\}\right] = \{(\boldsymbol{\xi}, v) \in \Xi \times V \mid v = y_0 + \boldsymbol{\xi} \cdot \boldsymbol{\zeta}_0\}. \tag{8}$$

That is the image under ψ of hyperplanes in $\mathcal{H}_1$ which intersect $(\boldsymbol{\zeta}_0, y_0) \in \mathbb{R}^n \times \mathbb{R}$ is a hyperplane in $\Xi \times V$.

Refinement of S. Let $S \subseteq \mathbb{R}^n$ be the set in Proposition 1.

Suppose $\mathfrak{h} \in \mathcal{H}_1$. Let $\phi_{\mathfrak{h}} : \mathbb{R}^n \to \mathbb{R}$ be the unique function such that $(\boldsymbol{\zeta}, \phi_{\mathfrak{h}}(\boldsymbol{\zeta})) \in \mathfrak{h}$ for all $\boldsymbol{\zeta} \in \mathbb{R}^n$.

Let $P_{\boldsymbol{\zeta},\mathfrak{h}} = \{(\boldsymbol{\xi}, v) \in \Xi \times V \mid v = \phi_{\mathfrak{h}}(\boldsymbol{\zeta}) + \boldsymbol{\xi} \cdot \boldsymbol{\zeta}\}$. By Corollary 8 and because ψ is injective, all hyperplanes in $\psi^{-1}[P_{\boldsymbol{\zeta},\mathfrak{h}}]$ intersect the point $(\boldsymbol{\zeta}, \phi_{\mathfrak{h}}(\boldsymbol{\zeta}))$ in the domain.

Let $\mathcal{P}_{\mathfrak{h}} = \{P_{\boldsymbol{\zeta},\mathfrak{h}} \mid \boldsymbol{\zeta} \in S\}$.

For unique $\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{n+1} \in S$, consider $\bigcap_{i \in [n+1]} P_{\boldsymbol{\zeta}_i,\mathfrak{h}}$. It follows any hyperplane in $\psi^{-1}\left[\bigcap_{i \in [n+1]} P_{\boldsymbol{\zeta}_i,\mathfrak{h}}\right]$ intersects the points $\{(\boldsymbol{\zeta}_1, \phi_{\mathfrak{h}}(\boldsymbol{\zeta}_1)), \ldots, (\boldsymbol{\zeta}_{n+1}, \phi_{\mathfrak{h}}(\boldsymbol{\zeta}_{n+1}))\}$ where each $\boldsymbol{\zeta}_i \in S$. By Corollary 1, these points do not lie on a common $n-1$ dimensional affine subspace, so $\mathfrak{h}$ is the only hyperplane in the domain of f intersecting

$$\{(\boldsymbol{\zeta}_1, \phi_{\mathfrak{h}}(\boldsymbol{\zeta}_1)), \ldots, (\boldsymbol{\zeta}_{n+1}, \phi_{\mathfrak{h}}(\boldsymbol{\zeta}_{n+1}))\}.$$

It follows $\bigcap_{i \in [n+1]} P_{\boldsymbol{\zeta}_i,\mathfrak{h}} = \{\psi(\mathfrak{h})\}$. Since $\mu \circ \varphi^{-1}$ is atomless, $|\mu \circ \varphi^{-1}|\left(\bigcap_{i \in [n+1]} P_{\boldsymbol{\zeta}_i,\mathfrak{h}}\right) = 0$.

By Lemma 3, there are only countably many $P_{\boldsymbol{\zeta},\mathfrak{h}} \in \mathcal{P}_{\mathfrak{h}}$ such that $|\mu \circ \varphi^{-1}|\left(P_{\boldsymbol{\zeta},\mathfrak{h}}\right) > 0$.

Define $S_{\mathfrak{h}} \coloneqq \{\boldsymbol{\zeta} \in S \mid |\mu \circ \varphi^{-1}|\left(P_{\boldsymbol{\zeta},\mathfrak{h}}\right) = 0\}$, so $S \setminus S_{\mathfrak{h}}$ is countable.

Let $\mathcal{H}_{supp}$ be the set of defining supporting hyperplanes of polyhedra in $\mathcal{C}$. Consider

$$S' := \bigcap_{\mathfrak{h}\in\mathcal{H}_1\cap\mathcal{H}_{supp}} S_{\mathfrak{h}}.$$

Since $\mathcal{H}_{supp}$ is countable, $S \setminus S'$ is countable. Since $B \cap S$ is uncountable for all open balls $B \subseteq \mathbb{R}^n$, S' is dense in $\mathbb{R}^n$. Notice, whenever $\boldsymbol{\zeta} \in S'$ and $(\boldsymbol{\zeta}, y)$ is on a hyperplane in $\mathcal{H}_1 \cap \mathcal{H}_{supp}$,

$$|\mu \circ \varphi^{-1}| \left(\{(\boldsymbol{\xi}, v) \in \Xi \times V \mid v = y + \boldsymbol{\xi}\cdot\boldsymbol{\zeta}\}\right) = 0. \tag{9}$$

Vanishing integrals over line segments. Suppose $\boldsymbol{\zeta}_0 \in \mathbb{R}^n$, $y_1, y_2 \in \mathbb{R}$. Suppose $y_1 \leq y_2$.

By Equation 8, $L_{\boldsymbol{\zeta}_0}(y_1, y_2)$ is the image under ψ of hyperplanes in $\mathcal{H}_1$ which intersect the line segment between $(\boldsymbol{\zeta}_0, y_1)$ (exclusive) and $(\boldsymbol{\zeta}_0, y_2)$ (inclusive). Therefore,

$$\varphi^{-1}\left[L_{\boldsymbol{\zeta}_0}(y_1, y_2)\right] = \left\{(\boldsymbol{a}, b) \in \mathcal{S}^n \times \mathbb{R} \mid \exists y' \in (y_1, y_2] \;\; \boldsymbol{a}\cdot(\boldsymbol{\zeta}_0, y') = b,\; \boldsymbol{a}\cdot\boldsymbol{e}_{n+1} \neq 0\right\}.$$

If $\varphi(\boldsymbol{a}, b) = (\boldsymbol{\xi}, v)$, then

$$\frac{1}{\sqrt{1+\sum_{i\in[n]}\xi_i^2}} = \frac{1}{\sqrt{1+\sum_{i\in[n]}\frac{a_i^2}{a_{n+1}^2}}} = \frac{a_{n+1}}{\sqrt{\sum_{i\in[n+1]}a_i^2}} = a_{n+1} = \boldsymbol{a}\cdot\boldsymbol{e}_{n+1}.$$

Therefore, as $(\boldsymbol{a}, b)$ such that $\boldsymbol{a}\cdot\boldsymbol{e}_{n+1} = 0$ do not contribute to the integral,

$$\int_{L_{\boldsymbol{\zeta}_0}(y_1,y_2)} \frac{1}{\sqrt{1+\|\boldsymbol{\xi}\|^2}}\,\mathrm{d}\mu\circ\varphi^{-1}(\boldsymbol{\xi}, v) = \int_{\varphi^{-1}\left[L_{\boldsymbol{\zeta}_0}(y_1,y_2)\right]} \boldsymbol{a}\cdot\boldsymbol{e}_{n+1}\,\mathrm{d}\mu(\boldsymbol{a}, b)$$
$$= \int_{\{(\boldsymbol{a},b)\in\mathcal{S}^n\times\mathbb{R} \mid \exists y'\in(y_1,y_2] \;\; \boldsymbol{a}\cdot(\boldsymbol{\zeta}_0,y')=b\}} \boldsymbol{a}\cdot\boldsymbol{e}_{n+1}\,\mathrm{d}\mu(\boldsymbol{a}, b).$$

By Lemma 1, for $y \in \mathbb{R}$,

$$D_{\boldsymbol{e}_{n+1}^+} f(\boldsymbol{\zeta}_0, y) = \int_{\{(\boldsymbol{a},b)\in\mathcal{S}^n\times\mathbb{R} \mid \boldsymbol{a}\cdot(\boldsymbol{\zeta}_0,y)\geq b\}} \boldsymbol{a}\cdot\boldsymbol{e}_{n+1}\,\mathrm{d}\mu(\boldsymbol{a}, b).$$

By definition of $\mathcal{S}^n$, $y \mapsto \boldsymbol{a}\cdot(\boldsymbol{\zeta}_0, y)$ is a non-decreasing, continuous function on $\mathbb{R}$ for any fixed $\boldsymbol{a} \in \mathcal{S}^n$. Thus, $\boldsymbol{a}\cdot(\boldsymbol{\zeta}_0, y_2) \geq \boldsymbol{a}\cdot(\boldsymbol{\zeta}_0, y_1)$ for all $\boldsymbol{a} \in \mathcal{S}^n$. Further, $\boldsymbol{a}\cdot(\boldsymbol{\zeta}_0, y_2) \geq b$ and $\boldsymbol{a}\cdot(\boldsymbol{\zeta}_0, y_1) < b$ if and only if $\boldsymbol{a}\cdot(\boldsymbol{\zeta}_0, y') = b$ for some $y' \in (y_1, y_2]$. Therefore,

$$D_{\boldsymbol{e}_{n+1}^+} f(\boldsymbol{\zeta}_0, y_2) - D_{\boldsymbol{e}_{n+1}^+} f(\boldsymbol{\zeta}_0, y_1) = \int_{L_{\boldsymbol{\zeta}_0}(y_1,y_2)} \frac{1}{\sqrt{1+\|\boldsymbol{\xi}\|^2}}\,\mathrm{d}\mu\circ\varphi^{-1}(\boldsymbol{\xi}, v).$$

It follows whenever $D_{\boldsymbol{e}_{n+1}^+} f(\boldsymbol{\zeta}_0, y_1) = D_{\boldsymbol{e}_{n+1}^+} f(\boldsymbol{\zeta}_0, y_2)$,

$$\int_{L_{\boldsymbol{\zeta}_0}(y_1,y_2)} \frac{1}{\sqrt{1+\|\boldsymbol{\xi}\|^2}}\,\mathrm{d}\mu\circ\varphi^{-1}(\boldsymbol{\xi}, v) = 0. \tag{10}$$

Vanishing integrals over half-spaces. Consider $\boldsymbol{\zeta}_0 \in S'$. Consider an interval $(y_0, \infty) \subseteq \mathbb{R}$.

For sets $E \subseteq \mathbb{R}^{n+1}$, let $\mathrm{ri}_{\boldsymbol{\zeta}_0}(E)$ denote the relative interior of $E \cap (\{\boldsymbol{\zeta}_0\} \times (-\infty, \infty))$ with respect to $\{\boldsymbol{\zeta}_0\} \times (-\infty, \infty)$. Then, define

$$J := \left\{ y \in (y_0, \infty) \mid (\boldsymbol{\zeta}_0, y) \in \bigcup_{C\in\mathcal{C}} \mathrm{ri}_{\boldsymbol{\zeta}_0}(C) \right\}.$$

It follows J is open. Then, there are countably many $q_i, r_i \in \mathbb{R} \cup \{\pm\infty\}$ such that $J = \bigcup_{i\in\mathbb{N}}(q_i, r_i)$, the intervals pairwise disjoint.

Moreover, $D_{\boldsymbol{e}_{n+1}^+} f(\boldsymbol{\zeta}, y)$ is constant on $\mathrm{ri}_{\boldsymbol{\zeta}_0}(C)$ for every $C \in \mathcal{C}$. As locally constant functions are constant on connected components, $D_{\boldsymbol{e}_{n+1}^+} f(\boldsymbol{\zeta}, y)$ is constant on $\{\boldsymbol{\zeta}_0\} \times (q_i, r_i)$ for all $i \in \mathbb{N}$. By Equation 10, for all $m \in \mathbb{N}$,

$$\int_{L_{\boldsymbol{\zeta}_0}\left(q_i+\frac{1}{m}, r_i-\frac{1}{m}\right)} \frac{1}{\sqrt{1+\|\boldsymbol{\xi}\|^2}} \,\mathrm{d}\mu \circ \varphi^{-1}(\boldsymbol{\xi}, v) = 0. \tag{11}$$

Thus, define $E_m \coloneqq \bigcup_{i\in\mathbb{N}} L_{\boldsymbol{\zeta}_0}\left(q_i + \frac{1}{m}, r_i - \frac{1}{m}\right)$ for $m \in \mathbb{N}$. By construction, this is a disjoint union. Therefore, by Equation 11, for all $m \in \mathbb{N}$,

$$\int_{E_m} \frac{1}{\sqrt{1+\|\boldsymbol{\xi}\|^2}} \,\mathrm{d}\mu \circ \varphi^{-1}(\boldsymbol{\xi}, v) = 0.$$

To extend the integral over all of $\{(\boldsymbol{\xi}, v) \in \Xi \times V \mid v > y_0 + \boldsymbol{\xi} \cdot \boldsymbol{\zeta}_0\}$, we must also address integrating over $\overline{L}_{\zeta_0}(y_1)$ for y_1 where $(\boldsymbol{\zeta}_0, y_1)$ lies on a boundary of an affine piece.

We will show $\left|\mu \circ \varphi^{-1}\right| \left(\bigcup_{b\in(y_0,\infty)\setminus J} \overline{L}_{\boldsymbol{\zeta}_0}(b)\right) = 0$.

Consider $C \in \mathcal{C}$. Suppose $y_1 \in \mathbb{R}$ is such that $(\boldsymbol{\zeta}_0, y_1) \in C$ and $(\boldsymbol{\zeta}_0, y_1)$ is not on a defining supporting hyperplane of C in $\mathcal{H}_1 \cap \mathcal{H}_{supp}$. In particular, either $(\boldsymbol{\zeta}_0, y_1)$ is on the interior of C or it lies only on defining supporting hyperplanes of C with normal vector orthogonal to $\boldsymbol{e}_{n+1}$. As C has only finitely many defining supporting hyperplanes, it follows there is $\delta > 0$ such that $(\boldsymbol{\zeta}_0, y_1 + \epsilon) \in C$ whenever $|\epsilon| < \delta$. Therefore, $(\boldsymbol{\zeta}_0, y_1) \in \mathrm{ri}_{\boldsymbol{\zeta}_0}(C)$.

Thus, $(y_0, \infty) \setminus J \subseteq \left\{y \in \mathbb{R} \mid (\boldsymbol{\zeta}_0, y) \in \bigcup_{\mathfrak{h}\in\mathcal{H}_1\cap\mathcal{H}_{supp}} \mathfrak{h}\right\}$. Further, for every $\mathfrak{h} \in \mathcal{H}_1 \cap \mathcal{H}_{supp}$, $|\mathfrak{h} \cap (\{\boldsymbol{\zeta}_0\} \times (y_0, \infty))| \le 1$. Therefore, as $\mathcal{H}_{supp}$ is countable, $(y_0, \infty) \setminus J$ is countable.

Suppose $b_0 \in (y_0, \infty) \setminus J$. Then, $(\boldsymbol{\zeta}_0, b_0) \in \bigcup_{\mathfrak{h}\in\mathcal{H}_1\cap\mathcal{H}_{supp}} \mathfrak{h}$. As $\boldsymbol{\zeta}_0 \in S'$, by Equation 9,

$$\left|\mu \circ \varphi^{-1}\right| \left(\overline{L}_{\boldsymbol{\zeta}_0}(b_0)\right) = \left|\mu \circ \varphi^{-1}\right| \left(\{(\boldsymbol{\xi}, v) \in \Xi \times V \mid v = b_0 + \boldsymbol{\xi} \cdot \boldsymbol{\zeta}_0\}\right) = 0. \tag{12}$$

Thus, as $(y_0, \infty) \setminus J$ is countable, $\left|\mu \circ \varphi^{-1}\right| \left(\bigcup_{b\in(y_0,\infty)\setminus J} \overline{L}_{\boldsymbol{\zeta}_0}(b)\right) = 0$.

It follows for all $m \in \mathbb{N}$,

$$\int_{E_m \cup \bigcup_{b\in(y_0,\infty)\setminus J} \overline{L}_{\boldsymbol{\zeta}_0}(b)} \frac{1}{\sqrt{1+\|\boldsymbol{\xi}\|^2}} \,\mathrm{d}\mu \circ \varphi^{-1}(\boldsymbol{\xi}, v) = 0.$$

Further, $E_m \cup \bigcup_{b\in(y_0,\infty)\setminus J} \overline{L}_{\boldsymbol{\zeta}_0}(b) \to \{(\boldsymbol{\xi}, v) \in \Xi \times V \mid v > y_0 + \boldsymbol{\xi} \cdot \boldsymbol{\zeta}_0\}$ as $m \to \infty$. By the Dominated Convergence Theorem, as $\mu \circ \varphi^{-1}$ is finite,

$$\int_{\{(\boldsymbol{\xi},v)\in\Xi\times V \mid v>y_0+\boldsymbol{\xi}\cdot\boldsymbol{\zeta}_0\}} \frac{1}{\sqrt{1+\|\boldsymbol{\xi}\|^2}} \,\mathrm{d}\mu \circ \varphi^{-1}(\boldsymbol{\xi}, v) = 0.$$

Similarly,

$$\int_{\{(\boldsymbol{\xi},v)\in\Xi\times V \mid v<y_0+\boldsymbol{\xi}\cdot\boldsymbol{\zeta}_0\}} \frac{1}{\sqrt{1+\|\boldsymbol{\xi}\|^2}} \,\mathrm{d}\mu \circ \varphi^{-1}(\boldsymbol{\xi}, v) = 0.$$

The equation $v = y_0 + \boldsymbol{\xi} \cdot \boldsymbol{\zeta}_0$ is equivalent to $(\boldsymbol{\zeta}_0, -1) \cdot (\boldsymbol{\xi}, v) = -y_0$. Therefore, for all $\boldsymbol{\zeta}_0 \in S'$ and $y_0 \in \mathbb{R}$, when considering an open half-space H in $\Xi \times V$ with a boundary defined by $(\boldsymbol{\zeta}_0, -1) \cdot (\boldsymbol{\xi}, v) = -y_0$, $\int_H \frac{1}{\sqrt{1+\|\boldsymbol{\xi}\|^2}} \,\mathrm{d}\mu \circ \varphi^{-1}(\boldsymbol{\xi}, v) = 0$.

Conclusion with Cramer–Wold and Radon–Nikodym. On the Borel sets of $\Xi \times V$, define the measure $\nu(E) = \int_E \frac{1}{\sqrt{1+\|\boldsymbol{\xi}\|^2}} \,\mathrm{d}\mu \circ \varphi^{-1}(\boldsymbol{\xi}, v)$. Given a Hahn decomposition of $\mu \circ \varphi^{-1}$ with positive set P and negative set N, $\nu(E) = \int_E (\chi_P - \chi_N) \frac{1}{\sqrt{1+\|\boldsymbol{\xi}\|^2}} \,\mathrm{d}|\mu \circ \varphi^{-1}|(\boldsymbol{\xi}, v)$.

By the previous part, if H is an open half-space with normal vector $(\boldsymbol{\zeta}_0, -1)$ with $\boldsymbol{\zeta}_0 \in S'$, then $\nu(H) = 0$.

Since S' is dense in $\mathbb{R}^n$, $\nu(H) = 0$ whenever the boundary of H is in a dense set of directions. By a careful inspection of the proof of the Cramer-Wold theorem, it follows the characteristic function of ν, $c_\nu(\boldsymbol{t}) \coloneqq \int_{\mathbb{R}^n} e^{i\boldsymbol{t}\cdot\boldsymbol{x}} \,\mathrm{d}\nu(\boldsymbol{x})$, is zero on a dense set of $\mathbb{R}^n$ [5, Equation 4]. By Dominated Convergence Theorem, in fact $c_\nu \equiv 0$. Since characteristic functions are unique, ν is the zero measure [13, Theorem 15.9].

However, the Radon-Nikodym derivative of a measure is unique up to almost everywhere. As 0 is a Radon-Nikodym derivative for the zero measure and $(\chi_P - \chi_N)\frac{1}{\sqrt{1+\|\boldsymbol{\xi}\|^2}}$ is never 0, it follows $|\mu \circ \varphi^{-1}|(\Xi \times V) = 0$.

Since φ is bijective between $\gamma^{-1}[\mathcal{H}_1]$ and $\Xi \times V$, the support of μ is contained in

$$(\mathcal{S}^n \times \mathbb{R}) \setminus \gamma^{-1}[\mathcal{H}_1] = \{\boldsymbol{a} \in \mathcal{S}^n \mid \boldsymbol{a} \cdot \boldsymbol{e}_{n+1} = 0\} \times \mathbb{R}.$$

By definition of $\mathcal{S}^n$, the support of μ is contained in a copy of $\mathcal{S}^{n-1} \times \mathbb{R}$ embedded into $\mathbb{S}^n \times \mathbb{R}$. That is, $f(\boldsymbol{x}) = \int_{\mathcal{S}^{n-1}\times\mathbb{R}} \sigma(\boldsymbol{a}\cdot\boldsymbol{x} - b) - \sigma(-b)\,\mathrm{d}\mu(\boldsymbol{a}, b)$. Moreover, $g : \mathbb{R}^n \to \mathbb{R}$ defined as

$$\begin{aligned} g(\boldsymbol{\zeta}) &\coloneqq \int_{\mathcal{S}^{n-1}\times\mathbb{R}} \sigma(\boldsymbol{\alpha}\cdot\boldsymbol{\zeta} - b) - \sigma(-b)\,\mathrm{d}\mu_{\mathcal{S}^{n-1}\times\mathbb{R}}(\boldsymbol{\alpha}, b) \\ &= \int_{\mathcal{S}^{n-1}\times\mathbb{R}} \sigma((\boldsymbol{\alpha}, 0)\cdot(\boldsymbol{\zeta}, 0) - b) - \sigma(-b)\,\mathrm{d}\mu_{\mathcal{S}^{n-1}\times\mathbb{R}}(\boldsymbol{\alpha}, b) = f(\boldsymbol{\zeta}, 0) \end{aligned}$$

is countably piecewise linear. By the inductive hypothesis, μ is the zero measure. □

We associate compactly supported measures in $\mathcal{M}(\mathbb{R}^{n+1} \times \mathbb{R})$ and measures in $\mathcal{M}(\mathbb{S}^n \times \mathbb{R})$ with those in $\mathcal{M}(\mathcal{S}^n \times \mathbb{R})$ in order to apply Theorem 1. Similar to this procedure, Ongie et al. in [15] decomposed measures in $\mathcal{M}(\mathbb{S}^n \times \mathbb{R})$ into even and odd components, where the odd component induced an affine function and the even component was unique.

Lemma 4. *Suppose τ is a compactly supported measure in $\mathcal{M}(\mathbb{R}^{n+1} \times \mathbb{R})$. Then, there exists $\mu \in \mathcal{M}(\mathbb{S}^n \times \mathbb{R})$ such that for all $\boldsymbol{x} \in \mathbb{R}^{n+1}$,*

$$\int_{\mathbb{R}^{n+1}\times\mathbb{R}} \sigma(\boldsymbol{a}\cdot\boldsymbol{x} - b) - \sigma(-b)\,\mathrm{d}\tau(\boldsymbol{a}, b) = \int_{\mathbb{S}^n\times\mathbb{R}} \sigma(\boldsymbol{a}\cdot\boldsymbol{x} - b) - \sigma(-b)\,\mathrm{d}\mu(\boldsymbol{a}, b).$$

Proof. Let $g : (\mathbb{R}^{n+1} \setminus \{\boldsymbol{0}\}) \times \mathbb{R} \to \mathbb{S}^n \times \mathbb{R}$ be defined as $g(\boldsymbol{a}, b) = \left(\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|}, \frac{b}{\|\boldsymbol{a}\|}\right)$. Let μ_1 be the Borel measure defined as $\mu_1(E) = \int_E \|\boldsymbol{a}\|\,\mathrm{d}\tau(\boldsymbol{a}, b)$. Since τ has compact support and is finite, $|\mu_1|(\mathbb{R}^{n+1} \times \mathbb{R}) < \infty$ and $\mu_1 \in \mathcal{M}(\mathbb{R}^{n+1} \times \mathbb{R})$. Let

$\mu \coloneqq \mu_1 \circ g^{-1}$. Then,

$$\begin{aligned}
&\int_{\mathbb{R}^{n+1}\times\mathbb{R}} \sigma\left(\boldsymbol{a}\cdot\boldsymbol{x}-b\right)-\sigma(-b)\,\mathrm{d}\tau\left(\boldsymbol{a},b\right)\\
&=\int_{(\mathbb{R}^{n+1}\setminus\{\boldsymbol{0}\})\times\mathbb{R}} \|\boldsymbol{a}\|\left(\sigma\left(\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|}\cdot x-\frac{b}{\|\boldsymbol{a}\|}\right)-\sigma\left(-\frac{b}{\|\boldsymbol{a}\|}\right)\right)\mathrm{d}\tau\left(\boldsymbol{a},b\right)\\
&\quad+\int_{\{\boldsymbol{0}\}\times\mathbb{R}} \sigma\left(-b\right)-\sigma(-b)\,\mathrm{d}\tau\left(\boldsymbol{a},b\right)\\
&=\int_{(\mathbb{R}^{n+1}\setminus\{\boldsymbol{0}\})\times\mathbb{R}} \sigma\left(\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|}\cdot\boldsymbol{x}-\frac{b}{\|\boldsymbol{a}\|}\right)-\sigma\left(-\frac{b}{\|\boldsymbol{a}\|}\right)\mathrm{d}\mu_1\left(\boldsymbol{a},b\right)+0\\
&=\int_{\mathbb{S}^n\times\mathbb{R}} \sigma\left(\boldsymbol{a}\cdot\boldsymbol{x}-b\right)-\sigma(-b)\,\mathrm{d}\mu\left(\boldsymbol{a},b\right).
\end{aligned}$$

□

Lemma 5. *Suppose $\tau \in \mathcal{M}\left(\mathbb{S}^n\times\mathbb{R}\right)$. Then, there exist $\mu \in \mathcal{M}\left(\mathcal{S}^n\times\mathbb{R}\right)$ and $\boldsymbol{a}_0 \in \mathbb{R}^{n+1}$ such that for all $\boldsymbol{x}\in\mathbb{R}^{n+1}$,*

$$\int_{\mathbb{S}^n\times\mathbb{R}} \sigma\left(\boldsymbol{a}\cdot\boldsymbol{x}-b\right)-\sigma(-b)\,\mathrm{d}\tau\left(\boldsymbol{a},b\right)=\int_{\mathcal{S}^n\times\mathbb{R}} \sigma\left(\boldsymbol{a}\cdot\boldsymbol{x}-b\right)-\sigma(-b)\,\mathrm{d}\mu\left(\boldsymbol{a},b\right)+\boldsymbol{a}_0\cdot\boldsymbol{x}.$$

Proof. Let $g\left(\boldsymbol{a},b\right)=\left(-\boldsymbol{a},-b\right)$ on $\mathcal{S}^n\times\mathbb{R}$. Let $\mu \coloneqq \tau+\tau\circ g^{-1}$ and $\boldsymbol{a}_0 \coloneqq -\int_{\mathcal{S}^n\times\mathbb{R}}\boldsymbol{a}\,\mathrm{d}\tau\circ g^{-1}\left(\boldsymbol{a},b\right)$. Note, $\sigma\left(-x\right)=\sigma\left(x\right)-x$. It follows

$$\begin{aligned}
&\int_{-\mathcal{S}^n\times\mathbb{R}} \sigma\left(\boldsymbol{a}\cdot\boldsymbol{x}-b\right)-\sigma(-b)\,\mathrm{d}\tau\left(\boldsymbol{a},b\right)\\
&=\int_{\mathcal{S}^n\times\mathbb{R}} \sigma\left(-\boldsymbol{a}\cdot\boldsymbol{x}+b\right)-\sigma(b)\,\mathrm{d}\tau\circ g^{-1}\left(\boldsymbol{a},b\right)\\
&=\int_{\mathcal{S}^n\times\mathbb{R}} \sigma\left(\boldsymbol{a}\cdot\boldsymbol{x}-b\right)-\left(\boldsymbol{a}\cdot\boldsymbol{x}-b\right)-\sigma(-b)-b\,\mathrm{d}\tau\circ g^{-1}\left(\boldsymbol{a},b\right)\\
&=\int_{\mathcal{S}^n\times\mathbb{R}} \sigma\left(\boldsymbol{a}\cdot\boldsymbol{x}-b\right)-\sigma(-b)\,\mathrm{d}\tau\circ g^{-1}\left(\boldsymbol{a},b\right)-\int_{\mathcal{S}^n\times\mathbb{R}}\boldsymbol{a}\cdot\boldsymbol{x}\,\mathrm{d}\tau\circ g^{-1}\left(\boldsymbol{a},b\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\int_{\mathbb{S}^n\times\mathbb{R}} \sigma\left(\boldsymbol{a}\cdot\boldsymbol{x}-b\right)-\sigma(-b)\,\mathrm{d}\tau\left(\boldsymbol{a},b\right)\\
&=\int_{\mathcal{S}^n\times\mathbb{R}} \sigma\left(\boldsymbol{a}\cdot\boldsymbol{x}-b\right)-\sigma(-b)\,\mathrm{d}\tau\left(\boldsymbol{a},b\right)+\int_{-\mathcal{S}^n\times\mathbb{R}} \sigma\left(\boldsymbol{a}\cdot\boldsymbol{x}-b\right)-\sigma(-b)\,\mathrm{d}\tau\left(\boldsymbol{a},b\right)\\
&=\int_{\mathcal{S}^n\times\mathbb{R}} \sigma\left(\boldsymbol{a}\cdot\boldsymbol{x}-b\right)-\sigma(-b)\,\mathrm{d}(\tau+\tau\circ g^{-1})\left(\boldsymbol{a},b\right)-\int_{\mathcal{S}^n\times\mathbb{R}}\boldsymbol{a}\,\mathrm{d}\tau\circ g^{-1}\left(\boldsymbol{a},b\right)\cdot\boldsymbol{x}\\
&=\int_{\mathcal{S}^n\times\mathbb{R}} \sigma\left(\boldsymbol{a}\cdot\boldsymbol{x}-b\right)-\sigma(-b)\,\mathrm{d}\mu\left(\boldsymbol{a},b\right)+\boldsymbol{a}_0\cdot\boldsymbol{x}.
\end{aligned}$$

□

To finish the proof of the conjecture (Corollary 3), it is necessary to split the measure into fully atomic and atomless parts and consider them separately. By first establishing point masses always induce nonaffineness even when dense, we can deduce the fully atomic and atomless components of the measure must both give rise to countably piecewise linear functions.

Lemma 6. *Let $\mu \in \mathcal{M}(\mathcal{S}^n \times \mathbb{R})$ and $f(\boldsymbol{x}) := \int_{\mathcal{S}^n \times \mathbb{R}} \sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b) - \sigma(-b)\, \mathrm{d}\mu(\boldsymbol{a}, b)$. Suppose $\mu(\{(\boldsymbol{c}, d)\}) \neq 0$. Then, $f(\boldsymbol{x})$ is not affine on every open ball in the domain of f intersecting $\mathfrak{h}_{\boldsymbol{c},d}$.*

Proof. Rotate the coordinate system of f such that $\boldsymbol{c} \cdot \boldsymbol{e}_{n+1} \neq 0$. Define $\Xi \times V$, ψ, and φ as in Theorem 1. Let $S \subseteq \mathbb{R}^n$ be the set in Proposition 1.

There exists a unique function $\phi : \mathbb{R}^n \to \mathbb{R}$ such that $(\boldsymbol{\zeta}, \phi(\boldsymbol{\zeta})) \in \mathfrak{h}_{\boldsymbol{c},d}$ for all $\boldsymbol{\zeta} \in \mathbb{R}^n$.

By way of contradiction, suppose f is affine on an open ball B_0 intersecting $\mathfrak{h}_{\boldsymbol{c},d}$. Then, there is an uncountable set $S' \subseteq S$ such that $\{(\boldsymbol{\zeta}, \phi(\boldsymbol{\zeta})) \mid \boldsymbol{\zeta} \in S'\} \subseteq B_0$.

For $\boldsymbol{\zeta} \in S'$, let $P_{\boldsymbol{\zeta}} = \{(\boldsymbol{\xi}, v) \in \Xi \times V \mid v = \phi(\boldsymbol{\zeta}) + \boldsymbol{\xi} \cdot \boldsymbol{\zeta}\}$. Now, let $\mathcal{P}'_{\boldsymbol{c},d} = \{P_{\boldsymbol{\zeta}} \setminus \{\psi(\mathfrak{h}_{\boldsymbol{c},d})\} \mid \boldsymbol{\zeta} \in S'\}$.

For unique $\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{n+1} \in S'$, consider $\bigcap_{i \in [n+1]} P_{\boldsymbol{\zeta}_i} \setminus \{\psi(\mathfrak{h}_{\boldsymbol{c},d})\}$.

It follows any hyperplane in $\psi^{-1}\left[\bigcap_{i \in [n+1]} P_{\boldsymbol{\zeta}_i} \setminus \{\psi(\mathfrak{h}_{\boldsymbol{c},d})\}\right]$ intersects the points $\{(\boldsymbol{\zeta}_1, \phi(\boldsymbol{\zeta}_1)), \ldots, (\boldsymbol{\zeta}_{n+1}, \phi(\boldsymbol{\zeta}_{n+1}))\}$ in the domain where each $\boldsymbol{\zeta}_i \in S$. By Corollary 1, these points do not lie on a common $n-1$ dimensional affine subspace, so $\mathfrak{h}_{\boldsymbol{c},d}$ is the only hyperplane intersecting $\{(\boldsymbol{\zeta}_1, \phi(\boldsymbol{\zeta}_1)), \ldots, (\boldsymbol{\zeta}_{n+1}, \phi(\boldsymbol{\zeta}_{n+1}))\}$. It follows $\bigcap_{i \in [n+1]} P_{\boldsymbol{\zeta}_i} \setminus \{\psi(\mathfrak{h}_{\boldsymbol{c},d})\} = \varnothing$ and $|\mu \circ \varphi^{-1}|\left(\bigcap_{i \in [n+1]} P_{\boldsymbol{\zeta}_i} \setminus \{\psi(\mathfrak{h}_{\boldsymbol{c},d})\}\right) = 0$.

Therefore, by Lemma 3, there are only countably many $P'_{\boldsymbol{\zeta}} \in \mathcal{P}'_{\boldsymbol{c},d}$ such that $|\mu \circ \varphi^{-1}|\left(P'_{\boldsymbol{\zeta}}\right) > 0$.

Since S' is uncountable, there is $\boldsymbol{\zeta}_0 \in S'$ such that $|\mu \circ \varphi^{-1}|\left(P_{\boldsymbol{\zeta}_0} \setminus \{\psi(\mathfrak{h}_{\boldsymbol{c},d})\}\right) = 0$.

As $(\boldsymbol{\zeta}_0, \phi(\boldsymbol{\zeta}_0)) \in B_0$, there is $\epsilon > 0$ such that for all $\delta < \epsilon$, f is affine on the line segment connecting $(\boldsymbol{\zeta}_0, \phi(\boldsymbol{\zeta}_0) - \delta)$ and $(\boldsymbol{\zeta}_0, \phi(\boldsymbol{\zeta}_0) + \delta)$.

By Equation 10 in Theorem 1, it follows

$$\int_{L_{\boldsymbol{\zeta}_0}(\phi(\boldsymbol{\zeta}_0)-\delta, \phi(\boldsymbol{\zeta}_0)+\delta)} \frac{1}{\sqrt{1 + \|\boldsymbol{\xi}\|^2}}\, \mathrm{d}\mu \circ \varphi^{-1}(\boldsymbol{\xi}, v) = 0.$$

Since this holds for all $\delta \in (0, \epsilon)$, by the Dominated Convergence Theorem,

$$\int_{\overline{L}_{\boldsymbol{\zeta}_0}(\phi(\boldsymbol{\zeta}_0))} \frac{1}{\sqrt{1 + \|\boldsymbol{\xi}\|^2}}\, \mathrm{d}\mu \circ \varphi^{-1}(\boldsymbol{\xi}, v) = 0. \tag{13}$$

Recall, $\overline{L}_{\boldsymbol{\zeta}_0}(\phi(\boldsymbol{\zeta}_0)) = \{(\boldsymbol{\xi}, v) \in \Xi \times V \mid v = \phi(\boldsymbol{\zeta}_0) + \boldsymbol{\xi} \cdot \boldsymbol{\zeta}_0\}$. By Equation 13,

$$\begin{aligned}
0 &= \int_{\overline{L}_{\boldsymbol{\zeta}_0}(\phi(\boldsymbol{\zeta}_0))} \frac{1}{\sqrt{1 + \|\boldsymbol{\xi}\|^2}}\, \mathrm{d}\mu \circ \varphi^{-1}(\boldsymbol{\xi}, v) \\
&= \int_{P_{\boldsymbol{\zeta}_0} \setminus \{\psi(\mathfrak{h}_{\boldsymbol{c},d})\}} \frac{1}{\sqrt{1 + \|\boldsymbol{\xi}\|^2}}\, \mathrm{d}\mu \circ \varphi^{-1}(\boldsymbol{\xi}, v) + \int_{\{\psi(\mathfrak{h}_{\boldsymbol{c},d})\}} \frac{1}{\sqrt{1 + \|\boldsymbol{\xi}\|^2}}\, \mathrm{d}\mu \circ \varphi^{-1}(\boldsymbol{\xi}, v) \\
&= 0 + \int_{\{(\boldsymbol{c},d)\}} \boldsymbol{a} \cdot \boldsymbol{e}_{n+1}\, \mathrm{d}\mu(\boldsymbol{a}, b) = \mu(\{(\boldsymbol{c}, d)\}) \cdot (\boldsymbol{c} \cdot \boldsymbol{e}_{n+1}).
\end{aligned}$$

This is a contradiction. □

Corollary 2. *Let $f : \mathbb{R}^{n+1} \to \mathbb{R}$ be a continuous countably piecewise linear function. Suppose there is a countable collection $\mathcal{C}$ of convex polyhedra covering $\mathbb{R}^{n+1}$ such that f is affine on each polyhedron and each polyhedron has non-empty interior. Suppose there exist $\mu \in \mathcal{M}(\mathbb{S}^n \times \mathbb{R})$ and $c_0 \in \mathbb{R}$ such that $f(\boldsymbol{x}) = \int_{\mathbb{S}^n \times \mathbb{R}} \sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b) - \sigma(-b)\, \mathrm{d}\mu(\boldsymbol{a}, b) + c_0$.*

Then, there are $r_0, r_{(\boldsymbol{c},d)} \in \mathbb{R}$ *and a countable set* $M \subseteq \mathcal{S}^n \times \mathbb{R}$ *such that* $f(\boldsymbol{x}) = r_0 + \sum_{(\boldsymbol{c},d)\in M} r_{(\boldsymbol{c},d)}\sigma(\boldsymbol{c}\cdot\boldsymbol{x} - d)$*.*

Proof. By Lemma 5, we can assume $f(\boldsymbol{x}) = \int_{\mathcal{S}^n\times\mathbb{R}} \sigma(\boldsymbol{a}\cdot\boldsymbol{x} - b) - \sigma(-b)\,\mathrm{d}\mu(\boldsymbol{a}, b) + \boldsymbol{a}_0\cdot\boldsymbol{x} + b_0$ for some $\mu \in \mathcal{M}(\mathcal{S}^n\times\mathbb{R})$, $\boldsymbol{a}_0 \in \mathbb{R}^{n+1}$, and $b_0 \in \mathbb{R}$. Decompose μ such that $\mu = \mu_C + \sum_{(\boldsymbol{c},d)\in M} r_{(\boldsymbol{c},d)}\delta_{(\boldsymbol{c},d)}$ where μ_C is atomless, M is a countable subset of $\mathcal{S}^n\times\mathbb{R}$, and $r_{(\boldsymbol{c},d)} \in \mathbb{R}\setminus\{0\}$ for all $(\boldsymbol{c}, d)$.

Let $g : \mathbb{R}^{n+1} \to \mathbb{R}$ be

$$\begin{aligned} g(\boldsymbol{x}) &\coloneqq \int_{\mathcal{S}^n\times\mathbb{R}} \sigma(\boldsymbol{a}\cdot\boldsymbol{x} - b) - \sigma(-b)\,\mathrm{d}\left(\sum_{(\boldsymbol{c},d)\in M} r_{(\boldsymbol{c},d)}\delta_{(\boldsymbol{c},d)}\right)(\boldsymbol{a}, b) + \boldsymbol{a}_0\cdot\boldsymbol{x} + b_0 \\ &= \sum_{(\boldsymbol{c},d)\in M} r_{(\boldsymbol{c},d)}\left(\sigma(\boldsymbol{c}\cdot\boldsymbol{x} - d) - \sigma(-d)\right) + \boldsymbol{a}_0\cdot\boldsymbol{x} + b_0. \end{aligned}$$

Then, g is certainly affine outside of $\bigcup_{(\boldsymbol{c},d)\in M}\mathfrak{h}_{\boldsymbol{c},d}$. By Lemma 6, for every $C \in \mathcal{C}$ and every $(\boldsymbol{c}, d) \in M$, $(\operatorname{int} C)\cap\mathfrak{h}_{\boldsymbol{c},d} = \varnothing$. Therefore, for every $C \in \mathcal{C}$, g is affine on C, because g is continuous and $\overline{\operatorname{int} C} = C$. Thus, the cover $\mathcal{C}$ shows g is countably piecewise linear.

It follows $\int_{\mathcal{S}^n\times\mathbb{R}} \sigma(\boldsymbol{a}\cdot\boldsymbol{x} - b) - \sigma(-b)\,\mathrm{d}\mu_C(\boldsymbol{a}, b)$ is also countably piecewise linear. By Theorem 1, μ_C is in fact the zero measure.

Then, $f(\boldsymbol{x}) = g(\boldsymbol{x})$. Note, $\boldsymbol{a}_0\cdot\boldsymbol{x} = \sigma(\boldsymbol{a}_0\cdot\boldsymbol{x}) - \sigma(-\boldsymbol{a}_0\cdot\boldsymbol{x})$. Thus,

$$f(\boldsymbol{x}) = \left(b_0 - \sum_{(\boldsymbol{c},d)\in M} r_{(\boldsymbol{c},d)}\sigma(-d)\right) + \sigma(\boldsymbol{a}_0\cdot\boldsymbol{x}) - \sigma(-\boldsymbol{a}_0\cdot\boldsymbol{x}) + \sum_{(\boldsymbol{c},d)\in M} r_{(\boldsymbol{c},d)}\sigma(\boldsymbol{c}\cdot\boldsymbol{x} - d).$$

□

Corollary 3. *Let* $f : \mathbb{R}^{n+1} \to \mathbb{R}$ *be a continuous finitely piecewise linear function. If there exist* $\mu \in \mathcal{M}(\mathbb{S}^n\times\mathbb{R})$ *and* $c_0 \in \mathbb{R}$ *such that* $f(\boldsymbol{x}) = \int_{\mathbb{S}^n\times\mathbb{R}} \sigma(\boldsymbol{a}\cdot\boldsymbol{x} - b) - \sigma(-b)\,\mathrm{d}\mu(\boldsymbol{a}, b) + c_0$*, then* f *is representable as a finite-width network as in Equation* 1*.*

Proof. Since f is *finitely* piecewise linear, by [9], there exists a finite collection $\mathcal{C}$ of convex polyhedra with non-empty interior covering $\mathbb{R}^{n+1}$ such that f is affine on each. By Corollary 2, there are $r_0 \in \mathbb{R}$, $r_{(\boldsymbol{c},d)} \in \mathbb{R}\setminus\{0\}$, and a countable set $M \subseteq \mathcal{S}^n\times\mathbb{R}$, such that $f(\boldsymbol{x}) = r_0 + \sum_{(\boldsymbol{c},d)\in M} r_{(\boldsymbol{c},d)}\sigma(\boldsymbol{c}\cdot\boldsymbol{x} - d)$. As f will have a boundary at $\mathfrak{h}_{\boldsymbol{c},d}$ for all $(\boldsymbol{c}, d) \in M$, M is a finite set. □

Corollary 4. *Let* $f(\boldsymbol{x}) = \int_{\mathbb{S}^n\times\mathbb{R}} \sigma(\boldsymbol{a}\cdot\boldsymbol{x} - b) - \sigma(-b)\,\mathrm{d}\mu(\boldsymbol{a}, b) + c_0$ *with* $\mu \in \mathcal{M}(\mathbb{S}^n\times\mathbb{R})$ *and* $c_0 \in \mathbb{R}$*. If* $f \not\equiv 0$*, then* f *is not a compactly supported finitely piecewise linear function.*

Proof. Suppose $f \not\equiv 0$. By Corollary 3, there are $r_0 \in \mathbb{R}$, $r_{(\boldsymbol{c},d)} \in \mathbb{R}\setminus\{0\}$, and a finite set $M \subseteq \mathcal{S}^n\times\mathbb{R}$ such that $f(\boldsymbol{x}) = r_0 + \sum_{(\boldsymbol{c},d)\in M} r_{(\boldsymbol{c},d)}\sigma(\boldsymbol{c}\cdot\boldsymbol{x} - d)$. If M is empty, $f(\boldsymbol{x}) = r_0 \neq 0$. Otherwise, f will not be affine along $\mathfrak{h}_{\boldsymbol{c},d}$ for some $(\boldsymbol{c}, d) \in M$. Since $n \geq 2$, $\mathfrak{h}_{\boldsymbol{c},d}$ will extend infinitely and f cannot be compactly supported. □

References

[1] Raman Arora, Amitabh Basu, Poorya Mianjy, and Anirbit Mukherjee, *Understanding Deep Neural Networks with Rectified Linear Units*, Preprint, `arXiv:1611.01491`, 2018.

[2] Francis Bach, *Breaking the curse of dimensionality with convex neutral networks*, J. Mach. Learn. Res. **18** (2017), Paper No. 19, 53. MR3634886

[3] Peter L Bartlett, *For Valid Generalization the Size of the Weights is More Important than the Size of the Network*, Advances in neural information processing systems, 1996, pp. 7.

[4] Emmanuel J. Candès, *Harmonic analysis of neural networks*, Appl. Comput. Harmon. Anal. **6** (1999), no. 2, 197–218, DOI 10.1006/acha.1998.0248. MR1676767

[5] H. Cramér and H. Wold, *Some Theorems on Distribution Functions*, J. London Math. Soc. **11** (1936), no. 4, 290–294, DOI 10.1112/jlms/s1-11.4.290. MR1574927

[6] G. Cybenko, *Approximation by superpositions of a sigmoidal function*, Math. Control Signals Systems **2** (1989), no. 4, 303–314, DOI 10.1007/BF02551274. MR1015670

[7] Weinan E, Chao Ma, and Lei Wu, *A priori estimates of the population risk for two-layer neural networks*, Commun. Math. Sci. **17** (2019), no. 5, 1407–1425, DOI 10.4310/CMS.2019.v17.n5.a11. MR4044196

[8] Weinan E. and Stephan Wojtowytsch, *Representation formulas and pointwise properties for Barron functions*, Calc. Var. Partial Differential Equations **61** (2022), no. 2, Paper No. 46, 37, DOI 10.1007/s00526-021-02156-6. MR4375792

[9] V. V. Gorokhovik, *Geometrical and analytical characteristic properties of piecewise affine mappings*, Preprint, `arXiv:1111.1389`, 2011.

[10] Paul R. Halmos, *Large intersections of large sets*, Amer. Math. Monthly **99** (1992), no. 4, 307–312, DOI 10.2307/2324896. MR1157221

[11] Juncai He, Lin Li, Jinchao Xu, and Chunyue Zheng, *Relu deep neural networks and linear finite elements*, J. Comput. Math. **38** (2020), no. 3, 502–527, DOI 10.4208/jcm.1901-m2018-0160. MR4087799

[12] Steve Jackson and R. Daniel Mauldin, *On the σ-class generated by open balls*, Math. Proc. Cambridge Philos. Soc. **127** (1999), no. 1, 99–108, DOI 10.1017/S0305004199003552. MR1692499

[13] Achim Klenke, *Characteristic Functions and the Central Limit Theorem*, Probability Theory: A Comprehensive Course, 2020, pp. 327–366.

[14] S. Kostadinova, S. Pilipović, K. Saneva, and J. Vindas, *The ridgelet transform of distributions*, Integral Transforms Spec. Funct. **25** (2014), no. 5, 344–358, DOI 10.1080/10652469.2013.853057. MR3172048

[15] Greg Ongie, Rebecca Willett, Daniel Soudry, and Nathan Srebro, *A function space view of bounded norm infinite width ReLU nets: The multivariate case*, Preprint, `arXiv:1910.01635`, 2019.

[16] Elisa Oostwal, Michiel Straat, and Michael Biehl, *Hidden unit specialization in layered neural networks: ReLU vs. sigmoidal activation*, Phys. A **564** (2021), Paper No. 125517, 14, DOI 10.1016/j.physa.2020.125517. MR4171794

[17] Pedro Savarese, Itay Evron, Daniel Soudry, and Nathan Srebro, *How do infinite width bounded norm networks look in function space?*, Proceedings of the Thirty-Second Conference on Learning Theory, PMLR 99, pp. 2667–2690, 2019.

[18] Sho Sonoda and Noboru Murata, *Neural network with unbounded activation functions is universal approximator*, Appl. Comput. Harmon. Anal. **43** (2017), no. 2, 233–268, DOI 10.1016/j.acha.2015.12.005. MR3668038

[19] J. v. Neumann, *Ein System algebraisch unabhängiger Zahlen* (German), Math. Ann. **99** (1928), no. 1, 134–141, DOI 10.1007/BF01459089. MR1512442

Department of Mathematics, Iowa State University, Ames, Iowa 50011
Email address: `smmcc@iastate.edu`