Neural network with unbounded activation functions is universal approximator

Sho Sonoda [*], Noboru Murata

Faculty of Science and Engineering, Waseda University, 3-4-1 Okubo Shinjuku-ku, Tokyo, 169-8555, Japan

A B S T R A C T

This paper presents an investigation of the approximation property of neural networks with unbounded activation functions, such as the rectified linear unit (ReLU), which is the new de-facto standard of deep learning. The ReLU network can be analyzed by the ridgelet transform with respect to Lizorkin distributions. By showing three reconstruction formulas by using the Fourier slice theorem, the Radon transform, and Parseval's relation, it is shown that a neural network with unbounded activation functions still satisfies the universal approximation property. As an additional consequence, the ridgelet transform, or the backprojection filter in the Radon domain, is what the network learns after backpropagation. Subject to a constructive admissibility condition, the trained network can be obtained by simply discretizing the ridgelet transform, without backpropagation. Numerical examples not only support the consistency of the admissibility condition but also imply that some non-admissible cases result in low-pass filtering.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Consider approximating a function $f : \mathbb{R}^m \to \mathbb{C}$ by the neural network g_J with an activation function $\eta : \mathbb{R} \to \mathbb{C}$

$$g_J(\mathbf{x}) = \frac{1}{J} \sum_j^J c_j \, \eta(\mathbf{a}_j \cdot \mathbf{x} - b_j), \quad (\mathbf{a}_j, b_j, c_j) \in \mathbb{R}^m \times \mathbb{R} \times \mathbb{C} \tag{1}$$

where we refer to $(\mathbf{a}_j, b_j)$ as a hidden parameter and c_j as an output parameter. Let $\mathbb{Y}^{m+1}$ denote the space of hidden parameters $\mathbb{R}^m \times \mathbb{R}$. The network g_J can be obtained by discretizing the *integral representation of the neural network*

---

* Corresponding author.
E-mail addresses: s.sonoda0110@toki.waseda.jp (S. Sonoda), noboru.murata@eb.waseda.ac.jp (N. Murata).

Table 1

Zoo of activation functions with which the corresponding neural network can approximate arbitrary functions in $L^1(\mathbb{R}^m)$ in the sense of pointwise convergence (§5.2) and in $L^2(\mathbb{R}^m)$ in the sense of mean convergence (§5.3). The third column indicates the space $\mathcal{W}(\mathbb{R})$ to which an activation function η belong (§6.1, 6.2).

Activation function	$\eta(z)$	$\mathcal{W}$
Unbounded functions		
Truncated power function	$z_+^k := \begin{cases} z^k & z > 0 \\ 0 & z \le 0 \end{cases}, \quad k \in \mathbb{N}_0$	$\mathcal{S}_0'$
Rectified linear unit (ReLU)	z_+	$\mathcal{S}_0'$
Softplus function	$\sigma^{(-1)}(z) := \log(1 + e^z)$	$\mathcal{O}_M$
Bounded but not integrable functions		
Unit step function	z_+^0	$\mathcal{S}_0'$
(Standard) sigmoidal function	$\sigma(z) := (1 + e^{-z})^{-1}$	$\mathcal{O}_M$
Hyperbolic tangent function	$\tanh(z)$	$\mathcal{O}_M$
Bump functions		
(Gaussian) radial basis function	$\mathrm{G}(z) := (2\pi)^{-1/2} \exp\left(-z^2/2\right)$	$\mathcal{S}$
The first derivative of sigmoidal function	$\sigma'(z)$	$\mathcal{S}$
Dirac's δ	$\delta(z)$	$\mathcal{S}_0'$
Oscillatory functions		
The kth derivative of RBF	$\mathrm{G}^{(k)}(z)$	$\mathcal{S}$
The kth derivative of sigmoidal function	$\sigma^{(k)}(z)$	$\mathcal{S}$
The kth derivative of Dirac's δ	$\delta^{(k)}(z)$	$\mathcal{S}_0'$

$$g(\mathbf{x}) = \int_{\mathbb{Y}^{m+1}} \mathrm{T}(\mathbf{a}, b) \eta(\mathbf{a} \cdot \mathbf{x} - b) \mathrm{d}\mu(\mathbf{a}, b), \tag{2}$$

where $\mathrm{T} : \mathbb{Y}^{m+1} \to \mathbb{C}$ corresponds to a continuous version of the output parameter; μ denotes a measure on $\mathbb{Y}^{m+1}$. The right-hand side expression is known as the *dual ridgelet transform* of T with respect to η

$$\mathscr{R}_\eta^\dagger \mathrm{T}(\mathbf{x}) = \int_{\mathbb{Y}^{m+1}} \mathrm{T}(\mathbf{a}, b) \eta(\mathbf{a} \cdot \mathbf{x} - b) \frac{\mathrm{d}\mathbf{a}\mathrm{d}b}{\|\mathbf{a}\|}. \tag{3}$$

By substituting in $\mathrm{T}(\mathbf{a}, b)$ the *ridgelet transform* of f with respect to ψ

$$\mathscr{R}_\psi f(\mathbf{a}, b) := \int_{\mathbb{R}^m} f(\mathbf{x}) \overline{\psi(\mathbf{a} \cdot \mathbf{x} - b)} \|\mathbf{a}\| \mathrm{d}\mathbf{x}, \tag{4}$$

under some good conditions, namely the *admissibility* of (ψ, η) and some regularity of f, we can reconstruct f by

$$\mathscr{R}_\eta^\dagger \mathscr{R}_\psi f = f. \tag{5}$$

By discretizing the reconstruction formula, we can verify the approximation property of neural networks with the activation function η.

In this study, we investigate the approximation property of neural networks for the case in which η is a Lizorkin distribution, by extensively constructing the ridgelet transform with respect to Lizorkin distributions. The Lizorkin distribution space $\mathcal{S}_0'$ is such a large space that contains the *rectified linear unit* (ReLU) z_+, *truncated power functions* z_+^k, and other unbounded functions that have at most polynomial growth (but do not have polynomials as such). Table 1 and Fig. 1 give some examples of Lizorkin distributions.

Recall that the derivative of the ReLU z_+ is the step function z_+^0. Formally, the following suggestive formula

$$\int_{\mathbb{Y}^{m+1}} \mathrm{T}(\mathbf{a}, b) \eta'(\mathbf{a} \cdot \mathbf{x} - b) \frac{\mathrm{d}\mathbf{a}\mathrm{d}b}{\|\mathbf{a}\|} = \int_{\mathbb{Y}^{m+1}} \partial_b \mathrm{T}(\mathbf{a}, b) \eta(\mathbf{a} \cdot \mathbf{x} - b) \frac{\mathrm{d}\mathbf{a}\mathrm{d}b}{\|\mathbf{a}\|}, \tag{6}$$
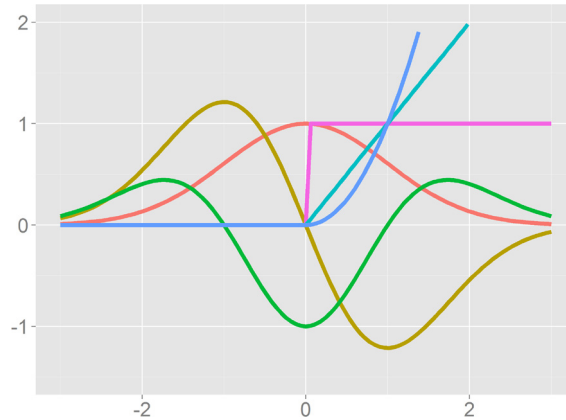
Fig. 1. Zoo of activation functions: the Gaussian $G(z)$ (red), the first derivative $G'(z)$ (yellow), the second derivative $G''(z)$ (green); a truncated power function z_+^2 (blue), the ReLU z_+ (sky blue), the unit step function z_+^0 (rose). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

holds, because the integral representation is a convolution in b. This formula suggests that once we have $T_{step}(\mathbf{a}, b)$ for the step function, which is implicitly known to exist based on some of our preceding studies [1,2], then we can formally obtain $T_{ReLU}(\mathbf{a}, b)$ for the ReLU by differentiating $T_{ReLU}(\mathbf{a}, b) = \partial_b T_{step}(\mathbf{a}, b)$.

1.1. ReLU and other unbounded activation functions

The ReLU [3–6] became a new building block of *deep neural networks*, in the place of traditional bounded activation functions such as the sigmoidal function and the radial basis function (RBF). Compared with traditional units, a neural network with the ReLU is said [3,7–9,6] to learn faster because it has larger gradients that can alleviate the vanishing gradient [3], and perform more efficiently because it extracts sparser features. To date, these hypotheses have only been empirically verified without analytical evaluation.

It is worth noting that in approximation theory, it was already shown in the 1990s that neural networks with such unbounded activation functions have the universal approximation property. To be precise, if the activation function is *not* a polynomial function, then the family of all neural networks is dense in some functional spaces such as $L^p(\mathbb{R}^m)$ and $C^0(\mathbb{R}^m)$. Mhaskar and Micchelli [10] seem to be the first to have shown such universality by using the B-spline. Later, Leshno et al. [11] reached a stronger claim by using functional analysis. Refer to Pinkus [12] for more details.

In this study, we initially work through the same statement by using harmonic analysis, or the ridgelet transform. One strength is that our results are very constructive. Therefore, we can construct what the network will learn during backpropagation. Note that for bounded cases this idea is already implicit in [13] and [2], and explicit in [14].

1.2. Integral representation of neural network and ridgelet transform

We use the *integral representation of neural networks* introduced by Murata [1]. As already mentioned, the integral representation corresponds to the dual ridgelet transform. In addition, the ridgelet transform corresponds to the composite of a wavelet transform after the Radon transform. Therefore, neural networks have a profound connection with harmonic analysis and tomography.

As Kůrková [15] noted, the idea of discretizing integral transforms to obtain an approximation is very old in approximation theory. As for neural networks, at first, Carroll and Dickinson [16] and Ito [13] regarded a neural network as a Radon transform [17]. Irie and Miyake [18], Funahashi [19], Jones [20], and Barron [21] used Fourier analysis to show the approximation property in a constructive way. Kůrková [15] applied

Barron's error bound to evaluate the complexity of neural networks. Refer to Kainen et al. [22] for more details.

In the late 1990s, Candès [23,24], Rubin [25], and Murata [1] independently proposed the so-called ridgelet transform, which has since been investigated by a number of authors [26–31].

1.3. Variations of ridgelet transform

A ridgelet transform $\mathscr{R}_\psi$, along with its reconstruction property, is determined by four classes of functions: domain $\mathcal{X}(\mathbb{R}^m)$, range $\mathcal{Y}(\mathbb{Y}^{m+1})$, ridgelet $\mathcal{Z}(\mathbb{R})$, and dual ridgelet $\mathcal{W}(\mathbb{R})$.

$$
\begin{array}{c}
\psi \in \mathcal{Z}(\mathbb{R}) \\
\overset{\mathscr{R}_\psi}{\nearrow} \\
\mathcal{X}(\mathbb{R}^m) \ni f \qquad\qquad\qquad T \in \mathcal{Y}(\mathbb{Y}^{m+1}) \\
\underset{\mathscr{R}_\eta^\dagger}{\swarrow} \\
\eta \in \mathcal{W}(\mathbb{R})
\end{array}
\tag{7}
$$

The following ladder relations by Schwartz [32] are fundamental for describing the variations of the ridgelet transform:

$$
\begin{array}{cccccc}
\text{(functions)} & \mathcal{D} \subset & \mathcal{S} \subset \mathcal{D}_{L^1} \subset \mathcal{D}_{L^p} \subset \mathcal{O}_\mathcal{M} \subset \mathcal{E} \\
& \cap & \cap \qquad \cap \qquad \cap \qquad \cap \qquad \cap \\
\text{(distributions)} & \mathcal{E}' \subset & \mathcal{O}'_\mathcal{C} \subset \mathcal{D}'_{L^1} \subset \mathcal{D}'_{L^p} \subset \mathcal{S}' \subset \mathcal{D}'
\end{array}
\tag{8}
$$

$$
\underbrace{\qquad\qquad\qquad\qquad}_{\text{integrable}} \quad \underbrace{\qquad\qquad\qquad\qquad}_{\text{not always bounded}}
$$

where the meaning of symbols are given below in Table 2.

The integral transform T by Murata [1] coincides with the case for $\mathcal{Z} \subset \mathcal{D}$ and $\mathcal{W} \subset \mathcal{E} \cap L^1$. Candès [23,24] proposed the "ridgelet transform" for $\mathcal{Z} = \mathcal{W} \subset \mathcal{S}$. Kostadinova et al. [30,31] defined the ridgelet transform for the Lizorkin distributions $\mathcal{X} = \mathcal{S}'_0$, which is the broadest domain ever known, at the cost of restricting the choice of ridgelet functions to the Lizorkin functions $\mathcal{W} = \mathcal{Z} = \mathcal{S}_0 \subset \mathcal{S}$.

1.4. Our goal

Although many researchers have investigated the ridgelet transform [26,29–31], in all the settings $\mathcal{Z}$ does *not* directly admit some fundamental activation functions, namely the sigmoidal function and the ReLU. One of the challenges we faced is to define the ridgelet transform for $\mathcal{W} = \mathcal{S}'_0$, which admits the sigmoidal function and the ReLU.

2. Preliminaries

2.1. Notations

Throughout this paper, we consider approximating $f : \mathbb{R}^m \to \mathbb{C}$ by a neural network g with hidden parameters $(\mathbf{a}, b)$. Following Kostadinova et al. [30,31], we denote by $\mathbb{Y}^{m+1} := \mathbb{R}^m \times \mathbb{R}$ the space of parameters $(\mathbf{a}, b)$. As already denoted, we symbolize the domain of a ridgelet transform as $\mathcal{X}(\mathbb{R}^m)$, the range as $\mathcal{Y}(\mathbb{Y}^{m+1})$, the space of ridgelets as $\mathcal{Z}(\mathbb{R})$, and the space of dual ridgelets as $\mathcal{W}(\mathbb{R})$.

Table 2
Classes of functions and distributions, and corresponding dual spaces.

Space	$\mathcal{A}(\mathbb{R}^k)$	Dual space	$\mathcal{A}'(\mathbb{R}^k)$
Polynomials of all degree	$\mathcal{P}(\mathbb{R}^k)$	–	
Smooth functions	$\mathcal{E}(\mathbb{R}^k)$	Compactly supported distributions	$\mathcal{E}'(\mathbb{R}^k)$
Rapidly decreasing functions	$\mathcal{S}(\mathbb{R}^k)$	Tempered distributions	$\mathcal{S}'(\mathbb{R}^k)$
Compactly supported smooth functions	$\mathcal{D}(\mathbb{R}^k)$	Schwartz distributions	$\mathcal{D}'(\mathbb{R}^k)$
L^p of Sobolev order ∞ $(1 \le p < \infty)$	$\mathcal{D}_{L^p}(\mathbb{R}^k)$	Schwartz dists. $(1/p + 1/q = 1)$	$\mathcal{D}'_{L^q}(\mathbb{R}^k)$
Completion of $\mathcal{D}(\mathbb{R}^k)$ in $\mathcal{D}_{L^\infty}(\mathbb{R}^k)$	$\dot{\mathcal{B}}(\mathbb{R}^k)$	Schwartz dists. $(p = 1)$	$\mathcal{D}'_{L^1}(\mathbb{R}^k)$
Slowly increasing functions	$\mathcal{O}_{\mathcal{M}}(\mathbb{R}^k)$	–	
–		Rapidly decreasing distributions	$\mathcal{O}'_{\mathcal{C}}(\mathbb{R}^k)$
Lizorkin functions	$\mathcal{S}_0(\mathbb{R}^k)$	Lizorkin distributions	$\mathcal{S}'_0(\mathbb{R}^k)$

We denote by $\mathbb{S}^{m-1}$ the $(m-1)$-sphere $\{\mathbf{u} \in \mathbb{R}^m \mid \|\mathbf{u}\| = 1\}$; by $\mathbb{R}_+$ the open half-line $\{\alpha \in \mathbb{R} \mid \alpha > 0\}$; by $\mathbb{H}$ the open half-space $\mathbb{R}_+ \times \mathbb{R}$. We denote by $\mathbb{N}$ and $\mathbb{N}_0$ the sets of natural numbers excluding 0 and including 0, respectively.

We denote by $\widetilde{\ \cdot\ }$ the reflection $\widetilde{f}(\mathbf{x}) := f(-\mathbf{x})$; by $\overline{\ \cdot\ }$ the complex conjugate; by $a \lesssim b$ that there exists a constant $C \ge 0$ such that $a \le Cb$.

2.2. Class of functions and distributions

Following Schwartz, we denote the classes of functions and distributions as in Table 2. For Schwartz's distributions, we refer to Schwartz [32] and Trèves [33]; for Lebesgue spaces, Rudin [34], Brezis [35] and Yosida [36]; for Lizorkin distributions, Yuan et al. [37] and Holschneider [38].

The space $\mathcal{S}_0(\mathbb{R}^k)$ of Lizorkin functions is a closed subspace of $\mathcal{S}(\mathbb{R}^k)$ that consists of elements such that all moments vanish. That is, $\mathcal{S}_0(\mathbb{R}^k) := \{\phi \in \mathcal{S}(\mathbb{R}^k) \mid \int_{\mathbb{R}^k} \mathbf{x}^{\boldsymbol{\alpha}} \phi(\mathbf{x}) \mathrm{d}\mathbf{x} = 0 \text{ for any } \boldsymbol{\alpha} \in \mathbb{N}_0^k\}$. The dual space $\mathcal{S}'_0(\mathbb{R}^k)$, known as the Lizorkin distribution space, is homeomorphic to the quotient space of $\mathcal{S}'(\mathbb{R}^k)$ by the space of all polynomials $\mathcal{P}(\mathbb{R}^k)$. That is, $\mathcal{S}'_0(\mathbb{R}^k) \cong \mathcal{S}'(\mathbb{R}^k)/\mathcal{P}(\mathbb{R}^k)$. Refer to Yuan et al. [37, Prop. 8.1] for more details. In this work we identify and treat every polynomial as zero in the Lizorkin distribution. That is, for $p \in \mathcal{S}'(\mathbb{R}^k)$, if $p \in \mathcal{P}(\mathbb{R}^k)$ then $p \equiv 0$ in $\mathcal{S}'_0(\mathbb{R}^k)$.

For $\mathbb{S}^{m-1}$, we work on the two subspaces $\mathcal{D}(\mathbb{S}^{m-1}) \subset \mathcal{D}(\mathbb{R}^m)$ and $\mathcal{E}'(\mathbb{S}^{m-1}) \subset \mathcal{E}'(\mathbb{R}^m)$. In addition, we identify $\mathcal{D} = \mathcal{S} = \mathcal{O}_{\mathcal{M}} = \mathcal{E}$ and $\mathcal{E}' = \mathcal{O}'_{\mathcal{C}} = \mathcal{S}' = \mathcal{D}'_{L^p} = \mathcal{D}'$.

For $\mathbb{H}$, let $\mathcal{E}(\mathbb{H}) \subset \mathcal{E}(\mathbb{R}^2)$ and $\mathcal{D}(\mathbb{H}) \subset \mathcal{D}(\mathbb{R}^2)$. For $\mathrm{T} \in \mathcal{E}(\mathbb{H})$, write

$$\mathrm{D}_{s,t}^{k,\ell}\mathrm{T}(\alpha, \beta) := (\alpha + 1/\alpha)^s (1 + \beta^2)^{t/2} \partial_\alpha^k \partial_\beta^\ell \mathrm{T}(\alpha, \beta), \quad s, t, k, \ell \in \mathbb{N}_0. \tag{9}$$

The space $\mathcal{S}(\mathbb{H})$ consists of $\mathrm{T} \in \mathcal{E}(\mathbb{H})$ such that for any $s, t, k, \ell \in \mathbb{N}_0$, the seminorm below is finite

$$\sup_{(\alpha,\beta)\in\mathbb{H}} \left|\mathrm{D}_{s,t}^{k,\ell}\mathrm{T}(\alpha,\beta)\right| < \infty. \tag{10}$$

The space $\mathcal{O}_{\mathcal{M}}(\mathbb{H})$ consists of $\mathrm{T} \in \mathcal{E}(\mathbb{H})$ such that for any $k, \ell \in \mathbb{N}_0$ there exist $s, t \in \mathbb{N}_0$ such that

$$\left|\mathrm{D}_{0,0}^{k,\ell}\mathrm{T}(\alpha,\beta)\right| \lesssim (\alpha + 1/\alpha)^s (1 + \beta^2)^{t/2}. \tag{11}$$

The space $\mathcal{D}'(\mathbb{H})$ consists of all bounded linear functionals Φ on $\mathcal{D}(\mathbb{H})$ such that for every compact set $\mathrm{K} \subset \mathbb{H}$, there exists $N \in \mathbb{N}_0$ such that

$$\left| \int_{\mathrm{K}} \mathrm{T}(\alpha, \beta) \Phi(\alpha, \beta) \frac{\mathrm{d}\alpha \mathrm{d}\beta}{\alpha} \right| \lesssim \sum_{k,\ell \le N} \sup_{(\alpha,\beta)\in\mathbb{H}} |\mathrm{D}_{0,0}^{k,\ell}\mathrm{T}(\alpha,\beta)|, \quad \forall \mathrm{T} \in \mathcal{D}(\mathrm{K}) \tag{12}$$

Table 3
Range of convolution (excerpt from Schwartz [32]).

Case	$\mathcal{A}_1$	$\mathcal{A}_2$	$\mathcal{A}_1 * \mathcal{A}_2$
Regularization	$\mathcal{D}$	$\mathcal{D}', \mathcal{D}'_{L^p}, \mathcal{E}'$	$\mathcal{E}, L^p, \mathcal{D}$
Compactly supported distribution	$\mathcal{E}'$	$\mathcal{E}', \mathcal{E}, \mathcal{D}'$	$\mathcal{E}', \mathcal{E}, \mathcal{D}'$
Regularization	$\mathcal{S}$	$\mathcal{S}, \mathcal{S}'$	$\mathcal{S}, \mathcal{O}_{\mathcal{M}}$
Schwartz convolutor	$\mathcal{O}'_{\mathcal{C}}$	$\mathcal{S}, \mathcal{O}'_{\mathcal{C}}, \mathcal{D}'_{L^p}, \mathcal{S}'$	$\mathcal{S}, \mathcal{O}'_{\mathcal{C}}, \mathcal{D}'_{L^p}, \mathcal{S}'$
Young's inequality	L^p	L^q	$L^r \ (1/r = 1/p + 1/q - 1)$
Young's inequality	$\mathcal{D}'_{L^p}$	$\mathcal{D}_{L^q}, \mathcal{D}'_{L^q}$	$\mathcal{D}'_{L^r} \ (1/r = 1/p + 1/q - 1)$

where the integral is understood as the action of Φ. The space $\mathcal{S}'(\mathbb{H})$ consists of $\Phi \in \mathcal{S}(\mathbb{H})$ for which there exists $N \in \mathbb{N}_0$ such that

$$\left| \int_{\mathbb{H}} \mathrm{T}(\alpha, \beta) \Phi(\alpha, \beta) \frac{\mathrm{d}\alpha \mathrm{d}\beta}{\alpha} \right| \lesssim \sum_{s,t,k,\ell \leq N} \sup_{(\alpha, \beta) \in \mathbb{H}} \left| \mathrm{D}_{s,t}^{k,\ell} \mathrm{T}(\alpha, \beta) \right|, \quad \forall \mathrm{T} \in \mathcal{S}(\mathbb{H}). \tag{13}$$

2.3. Convolution of distributions

Table 3 lists the convergent convolutions of distributions and their ranges by Schwartz [32].

In general a convolution of distributions may neither commute $\phi * \psi \neq \psi * \phi$ nor associate $\phi * (\psi * \eta) \neq (\phi * \psi) * \eta$. According to Schwartz [32, Ch. 6, Th. 7, Ch. 7, Th. 7], both $\mathcal{D}' * \mathcal{E}' * \mathcal{E}' * \cdots$ and $\mathcal{S}' * \mathcal{O}'_{\mathcal{C}} * \mathcal{O}'_{\mathcal{C}} * \cdots$ are commutative and associative.

2.4. Fourier analysis

The Fourier transform $\widehat{\cdot}$ of $f : \mathbb{R}^m \to \mathbb{C}$ and the inverse Fourier transform $\check{\cdot}$ of $F : \mathbb{R}^m \to \mathbb{C}$ are given by

$$\widehat{f}(\boldsymbol{\xi}) := \int_{\mathbb{R}^m} f(\mathbf{x}) e^{-i\mathbf{x} \cdot \boldsymbol{\xi}} \mathrm{d}\mathbf{x}, \quad \boldsymbol{\xi} \in \mathbb{R}^m \tag{14}$$

$$\check{F}(\mathbf{x}) := \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} F(\boldsymbol{\xi}) e^{i\mathbf{x} \cdot \boldsymbol{\xi}} \mathrm{d}\boldsymbol{\xi}, \quad \mathbf{x} \in \mathbb{R}^m. \tag{15}$$

The Hilbert transform $\mathcal{H}$ of $f : \mathbb{R} \to \mathbb{C}$ is given by

$$\mathcal{H}f(s) := \frac{i}{\pi} \mathrm{p.v.} \int_{-\infty}^{\infty} \frac{f(t)}{s - t} \mathrm{d}t, \quad s \in \mathbb{R} \tag{16}$$

where p.v. $\int_{-\infty}^{\infty}$ denotes the principal value. We set the coefficients above to satisfy

$$\widehat{\mathcal{H}f}(\omega) = \mathrm{sgn}\,\omega \cdot \widehat{f}(\omega), \tag{17}$$

$$\mathcal{H}^2 f(s) = f(s). \tag{18}$$

2.5. Radon transform

The Radon transform R of $f : \mathbb{R}^m \to \mathbb{C}$ and the dual Radon transform R^* of $\Phi : \mathbb{S}^{m-1} \times \mathbb{R} \to \mathbb{C}$ are given by

$$\mathrm{R}f(\mathbf{u}, p) := \int_{(\mathbb{R}\mathbf{u})^\perp} f(p\mathbf{u} + \mathbf{y})\mathrm{d}\mathbf{y}, \quad (\mathbf{u}, p) \in \mathbb{S}^{m-1} \times \mathbb{R} \tag{19}$$

$$\mathrm{R}^*\Phi(\mathbf{x}) := \int_{\mathbb{S}^{m-1}} \Phi(\mathbf{u}, \mathbf{u} \cdot \mathbf{x})\mathrm{d}\mathbf{u}, \quad \mathbf{x} \in \mathbb{R}^m \tag{20}$$

where $(\mathbb{R}\mathbf{u})^\perp := \{\mathbf{y} \in \mathbb{R}^m \mid \mathbf{y} \cdot \mathbf{u} = 0\}$ denotes the orthogonal complement of a line $\mathbb{R}\mathbf{u} \subset \mathbb{R}^m$; and $\mathrm{d}\mathbf{y}$ denotes the Lebesgue measure on $(\mathbb{R}\mathbf{u})^\perp$; and $\mathrm{d}\mathbf{u}$ denotes the surface measure on $\mathbb{S}^{m-1}$.

We use the following fundamental results [17,39] for $f \in L^1(\mathbb{R}^m)$ without proof: *Radon's inversion formula*

$$\mathrm{R}^*\Lambda^{m-1}\mathrm{R}f = 2(2\pi)^{m-1}f, \tag{21}$$

where the *backprojection filter* Λ^m is defined in (24); the *Fourier slice theorem*

$$\widehat{f}(\omega\mathbf{u}) = \int_{\mathbb{R}} \mathrm{R}f(\mathbf{u}, p)e^{-ip\omega}\mathrm{d}p, \quad (\mathbf{u}, \omega) \in \mathbb{S}^{m-1} \times \mathbb{R} \tag{22}$$

where the left-hand side is the m-dimensional Fourier transform, whereas the right-hand side is the one-dimensional Fourier transform of the Radon transform; and a corollary of *Fubini's theorem*

$$\int_{\mathbb{R}} \mathrm{R}f(\mathbf{u}, p)\mathrm{d}p = \int_{\mathbb{R}^m} f(\mathbf{x})\mathrm{d}\mathbf{x}, \quad \text{a.e. } \mathbf{u} \in \mathbb{S}^{m-1}. \tag{23}$$

2.6. Backprojection filter

For a function $\Phi(\mathbf{u}, p)$, we define the *backprojection filter* Λ^m as

$$\Lambda^m \Phi(\mathbf{u}, p) := \begin{cases} \partial_p^m \Phi(\mathbf{u}, p), & m \text{ even} \\ \mathcal{H}_p \partial_p^m \Phi(\mathbf{u}, p), & m \text{ odd.} \end{cases} \tag{24}$$

where $\mathcal{H}_p$ and ∂_p denote the Hilbert transform and the partial differentiation with respect to p, respectively. It is designed as a one-dimensional Fourier multiplier with respect to $p \to \omega$ such that

$$\widehat{\Lambda^m\Phi}(\mathbf{u}, \omega) = i^m |\omega|^m \widehat{\Phi}(\mathbf{u}, \omega). \tag{25}$$

3. Classical ridgelet transform

3.1. An overview

The *ridgelet transform* $\mathscr{R}_\psi f$ of $f : \mathbb{R}^m \to \mathbb{C}$ with respect to $\psi : \mathbb{R} \to \mathbb{C}$ is formally given by

$$\mathscr{R}_\psi f(\mathbf{a}, b) := \int_{\mathbb{R}^m} f(\mathbf{x})\overline{\psi(\mathbf{a} \cdot \mathbf{x} - b)}\|\mathbf{a}\|^s \mathrm{d}\mathbf{x}, \quad (\mathbf{a}, b) \in \mathbb{Y}^{m+1} \text{ and } s > 0. \tag{26}$$

The factor $|\mathbf{a}|^s$ is simply posed for technical convenience. After the next section we set $s = 1$, which simplifies some notations (e.g., Theorem 4.2). Murata [1] originally posed $s = 0$, which is suitable for the Euclidean formulation. Other authors such as Candès [24] used $s = 1/2$, Rubin [25] used $s = m$, and Kostadinova et al. [30] used $s = 1$.

When $f \in L^1(\mathbb{R}^m)$ and $\psi \in L^\infty(\mathbb{R})$, by using Hölder's inequality, the ridgelet transform is absolutely convergent at every $(\mathbf{a}, b) \in \mathbb{Y}^{m+1}$,

$$\int_{\mathbb{R}^m} \left| f(\mathbf{x}) \overline{\psi(\mathbf{a} \cdot \mathbf{x} - b)} \|\mathbf{a}\|^s \right| \mathrm{d}\mathbf{x} \leq \|f\|_{L^1(\mathbb{R}^m)} \cdot \|\psi\|_{L^\infty(\mathbb{R})} \cdot \|\mathbf{a}\|^s < \infty. \tag{27}$$

In particular when $s = 0$, the estimate is independent of $\mathbf{a}$ and thus $\mathscr{R}_\psi f \in L^\infty(\mathbb{Y}^{m+1})$. Furthermore, $\mathscr{R}$ is a bounded bilinear operator $L^1(\mathbb{R}^m) \times L^\infty(\mathbb{R}) \to L^\infty(\mathbb{Y}^{m+1})$.

The *dual ridgelet transform* $\mathscr{R}_\eta^\dagger \mathrm{T}$ of $\mathrm{T} : \mathbb{Y}^{m+1} \to \mathbb{C}$ with respect to $\eta : \mathbb{R} \to \mathbb{C}$ is formally given by

$$\mathscr{R}_\eta^\dagger \mathrm{T}(\mathbf{x}) := \int_{\mathbb{Y}^{m+1}} \mathrm{T}(\mathbf{a}, b) \eta(\mathbf{a} \cdot \mathbf{x} - b) \|\mathbf{a}\|^{-s} \mathrm{d}\mathbf{a} \mathrm{d}b, \quad \mathbf{x} \in \mathbb{R}^m. \tag{28}$$

The integral is absolutely convergent when $\eta \in L^\infty(\mathbb{R})$ and $\mathrm{T} \in L^1(\mathbb{Y}^{m+1}; \|\mathbf{a}\|^{-s} \mathrm{d}\mathbf{a} \mathrm{d}b)$ at every $\mathbf{x} \in \mathbb{R}^m$,

$$\int_{\mathbb{Y}^{m+1}} \left| \mathrm{T}(\mathbf{a}, b) \eta(\mathbf{a} \cdot \mathbf{x} - b) \right| \|\mathbf{a}\|^{-s} \mathrm{d}\mathbf{a} \mathrm{d}b \leq \|\mathrm{T}\|_{L^1(\mathbb{Y}^{m+1}; \|\mathbf{a}\|^{-s} \mathrm{d}\mathbf{a} \mathrm{d}b)} \cdot \|\eta\|_{L^\infty(\mathbb{R})} < \infty, \tag{29}$$

and thus $\mathscr{R}^\dagger$ is a bounded bilinear operator $L^1(\mathbb{Y}^{m+1}; \|\mathbf{a}\|^{-s} \mathrm{d}\mathbf{a} \mathrm{d}b) \times L^\infty(\mathbb{R}) \to L^\infty(\mathbb{R}^m)$.

Two functions ψ and η are said to be *admissible* when

$$K_{\psi,\eta} := (2\pi)^{m-1} \int_{-\infty}^{\infty} \frac{\overline{\widehat{\psi}(\zeta)} \widehat{\eta}(\zeta)}{|\zeta|^m} \mathrm{d}\zeta, \tag{30}$$

is finite and not zero. Provided that ψ, η, and f belong to some good classes, and ψ and η are admissible, then the *reconstruction formula*

$$\mathscr{R}_\eta^\dagger \mathscr{R}_\psi f = K_{\eta,\psi} f, \tag{31}$$

holds.

3.2. Ridgelet transform in other expressions

It is convenient to write the ridgelet transform in "polar" coordinates as

$$\mathscr{R}_\psi f(\mathbf{u}, \alpha, \beta) = \int_{\mathbb{R}^m} f(\mathbf{x}) \overline{\psi\left(\frac{\mathbf{u} \cdot \mathbf{x} - \beta}{\alpha}\right)} \frac{1}{\alpha^s} \mathrm{d}\mathbf{x}, \tag{32}$$

where "polar" variables are given by

$$\mathbf{u} := \mathbf{a}/\|\mathbf{a}\|, \quad \alpha := 1/\|\mathbf{a}\|, \quad \beta := b/\|\mathbf{a}\|. \tag{33}$$

Emphasizing the connection with wavelet analysis, we define the "radius" α as reciprocal. Provided there is no likelihood of confusion, we use the same symbol $\mathbb{Y}^{m+1}$ for the parameter space, regardless of whether it is parametrized by $(\mathbf{a}, b) \in \mathbb{R}^m \times \mathbb{R}$ or $(\mathbf{u}, \alpha, \beta) \in \mathbb{S}^{m-1} \times \mathbb{R}_+ \times \mathbb{R}$.

For a fixed $(\mathbf{u}, \alpha, \beta) \in \mathbb{Y}^{m+1}$, the ridgelet function

$$\psi_{\mathbf{u},\alpha,\beta}(\mathbf{x}) := \psi\left(\frac{\mathbf{u} \cdot \mathbf{x} - \beta}{\alpha}\right) \frac{1}{\alpha^s}, \quad \mathbf{x} \in \mathbb{R}^m \tag{34}$$

behaves as a constant function on $(\mathbb{R}\mathbf{u})^{\perp}$, and as a dilated and translated wavelet function on $\mathbb{R}\mathbf{u}$. That is, by using the orthogonal decomposition $\mathbf{x} = p\mathbf{u} + \mathbf{y}$ with $p \in \mathbb{R}$ and $\mathbf{y} \in (\mathbb{R}\mathbf{u})^{\perp}$,

$$\psi_{\mathbf{u},\alpha,\beta}(\mathbf{x}) = \psi\left(\frac{\mathbf{u}\cdot(p\mathbf{u}+\mathbf{y})-\beta}{\alpha}\right)\frac{1}{\alpha^s} = \psi\left(\frac{p-\beta}{\alpha}\right)\frac{1}{\alpha^s}\otimes 1(\mathbf{y}). \tag{35}$$

By using the decomposition above and Fubini's theorem, and assuming that the ridgelet transform is absolutely convergent, we have the following equivalent expressions

$$\mathscr{R}_{\psi}f(\mathbf{u},\alpha,\beta) = \int_{\mathbb{R}}\left(\int_{(\mathbb{R}\mathbf{u})^{\perp}} f(p\mathbf{u}+\mathbf{y})\mathrm{d}\mathbf{y}\right)\overline{\psi\left(\frac{p-\beta}{\alpha}\right)}\frac{1}{\alpha^s}\mathrm{d}p \tag{36}$$

$$= \int_{\mathbb{R}} \mathrm{R}f(\mathbf{u},p)\,\overline{\psi\left(\frac{p-\beta}{\alpha}\right)}\frac{1}{\alpha^s}\mathrm{d}p \tag{37}$$

$$= \int_{\mathbb{R}} \alpha^{1-s}\mathrm{R}f(\mathbf{u},\alpha z+\beta)\,\overline{\psi(z)}\mathrm{d}z \qquad\qquad \text{(weak form)} \tag{38}$$

$$= \left(\mathrm{R}f(\mathbf{u},\cdot)*\overline{\widetilde{\psi_{\alpha}}}\right)(\beta), \quad \psi_{\alpha}(p) := \psi\left(\frac{p}{\alpha}\right)\frac{1}{\alpha^s} \qquad \text{(convolution form)} \tag{39}$$

$$= \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{f}(\omega\mathbf{u})\overline{\widehat{\psi}(\alpha\omega)}\alpha^{1-s}e^{i\omega\beta}\mathrm{d}\omega, \qquad\qquad \text{(Fourier slice th. [30])} \tag{40}$$

where R denotes the Radon transform (19); the Fourier form follows by applying the identity $\mathcal{F}_{\omega}^{-1}\mathcal{F}_p = \mathrm{Id}$ to the convolution form. These reformulations reflect a well-known claim [28,30] that ridgelet analysis is wavelet analysis in the Radon domain.

3.3. Dual ridgelet transform in other expressions

Provided the dual ridgelet transform (28) is absolutely convergent, some changes of variables lead to other expressions.

$$\mathscr{R}_{\eta}^{\dagger}\mathrm{T}(\mathbf{x}) = \int_{\mathbb{R}^m}\int_{\mathbb{R}} \mathrm{T}(\mathbf{a},b)\eta(\mathbf{a}\cdot\mathbf{x}-b)\|\mathbf{a}\|^{-s}\mathrm{d}b\,\mathrm{d}\mathbf{a} \tag{41}$$

$$= \int_0^{\infty}\int_{\mathbb{S}^{m-1}}\int_{\mathbb{R}} \mathrm{T}(r\mathbf{u},b)\eta(r\mathbf{u}\cdot\mathbf{x}-b)\mathrm{d}b\,\mathrm{d}\mathbf{u}\,r^{m-s-1}\mathrm{d}r \tag{42}$$

$$= \int_{\mathbb{S}^{m-1}}\int_0^{\infty}\int_{\mathbb{R}} \mathrm{T}\left(\frac{\mathbf{u}}{\alpha},\frac{\beta}{\alpha}\right)\eta\left(\frac{\mathbf{u}\cdot\mathbf{x}-\beta}{\alpha}\right)\frac{\mathrm{d}\beta\mathrm{d}\alpha\mathrm{d}\mathbf{u}}{\alpha^{m-s+2}} \quad \text{(polar expression)} \tag{43}$$

$$= \int_{\mathbb{S}^{m-1}}\int_0^{\infty}\int_{\mathbb{R}} \mathrm{T}(\mathbf{u},\alpha,\mathbf{u}\cdot\mathbf{x}-\alpha z)\,\eta(z)\frac{\mathrm{d}z\mathrm{d}\alpha\mathrm{d}\mathbf{u}}{\alpha^{m-s+1}}, \quad \text{(weak form)} \tag{44}$$

where every integral is understood to be an iterated integral; the second equation follows by substituting $(r,\mathbf{u}) \leftarrow (\|\mathbf{a}\|, \mathbf{a}/\|\mathbf{a}\|)$ and using the coarea formula for polar coordinates; the third equation follows by substituting $(\alpha,\beta) \leftarrow (1/r,b/r)$ and using Fubini's theorem; in the fourth equation with a slight abuse of notation, we write $\mathrm{T}(\mathbf{u},\alpha,\beta) := \mathrm{T}(\mathbf{u}/\alpha,\beta/\alpha)$.

Table 4
Combinations of classes for which the ridgelet transform is well defined as a bilinear map. The first and third columns list domains $\mathcal{X}(\mathbb{R}^m)$ of f and $\mathcal{Z}(\mathbb{R})$ of ψ, respectively. The second column lists the range of the Radon transform $\mathrm{R}f(\mathbf{u}, p)$ for which we reused the same symbol $\mathcal{X}$ as it coincides. The fourth, fifth, and sixth columns list the range of the ridgelet transform with respect to β, (α, β), and $(\mathbf{u}, \alpha, \beta)$, respectively.

$f(\mathbf{x})$	$\mathrm{R}f(\mathbf{u}, p)$	$\psi(z)$	$\mathscr{R}_\psi f(\mathbf{u}, \alpha, \beta)$		
$\mathcal{X}(\mathbb{R}^m)$	$\mathcal{X}(\mathbb{S}^{m-1} \times \mathbb{R})$	$\mathcal{Z}(\mathbb{R})$	$\mathcal{B}(\mathbb{R})$	$\mathcal{A}(\mathbb{H})$	$\mathcal{Y}(\mathbb{Y}^{m+1})$
$\mathcal{D}$	$\mathcal{D}$	$\mathcal{D}'$	$\mathcal{E}$	$\mathcal{E}$	$\mathcal{E}$
$\mathcal{E}'$	$\mathcal{E}'$	$\mathcal{D}'$	$\mathcal{D}'$	$\mathcal{D}'$	$\mathcal{D}'$
$\mathcal{S}$	$\mathcal{S}$	$\mathcal{S}'$	$\mathcal{O}_\mathcal{M}$	$\mathcal{O}_\mathcal{M}$	$\mathcal{O}_\mathcal{M}$
$\mathcal{O}'_\mathcal{C}$	$\mathcal{O}'_\mathcal{C}$	$\mathcal{S}'$	$\mathcal{S}'$	$\mathcal{S}'$	$\mathcal{S}'$
L^1	L^1	$L^p \cap C^0$	$L^p \cap C^0$	$\mathcal{S}'$	$\mathcal{S}'$
$\mathcal{D}'_{L^1}$	$\mathcal{D}'_{L^1}$	$\mathcal{D}'_{L^p}$	$\mathcal{D}'_{L^p}$	$\mathcal{S}'$	$\mathcal{S}'$

Furthermore, write $\eta_\alpha(p) := \eta(p/\alpha)/\alpha^t$. Recall that the dual Radon transform R^* is given by (20) and the Mellin transform $\mathcal{M}$ [38] is given by $\mathcal{M}f(z) := \int_0^\infty f(\alpha)\alpha^{z-1}\mathrm{d}\alpha, \ z \in \mathbb{C}$. Then,

$$\mathscr{R}_\eta^\dagger \mathrm{T}(\mathbf{x}) = \mathrm{R}^* \left[\mathcal{M}[\mathrm{T}(\mathbf{u}, \alpha, \cdot) * \eta_\alpha](s + t - m - 1) \right](\mathbf{x}). \tag{45}$$

Note that the composition of the Mellin transform and the convolution is the dual wavelet transform [38]. Thus, the dual ridgelet transform is the composition of the dual Radon transform and the dual wavelet transform.

4. Ridgelet transform with respect to distributions

Using the weak expressions (38) and (44), we define the ridgelet transform with respect to distributions. Henceforth, we focus on the case for which the index s in (26) equals 1.

4.1. Definition and well-definedness

Definition 4.1 *(Ridgelet transform with respect to distributions).* The ridgelet transform $\mathscr{R}_\psi f$ of a function $f \in \mathcal{X}(\mathbb{R}^m)$ with respect to a distribution $\psi \in \mathcal{Z}(\mathbb{R})$ is given by

$$\mathscr{R}_\psi f(\mathbf{u}, \alpha, \beta) := \int_\mathbb{R} \mathrm{R}f(\mathbf{u}, \alpha z + \beta) \overline{\psi(z)}\mathrm{d}z, \quad (\mathbf{u}, \alpha, \beta) \in \mathbb{Y}^{m+1} \tag{46}$$

where $\int_\mathbb{R} \cdot \, \overline{\psi(z)}\mathrm{d}z$ is understood as the action of a distribution ψ.

Obviously, this "weak" definition coincides with the ordinary strong one when ψ coincides with a locally integrable function (L^1_{loc}). With a slight abuse of notation, the weak definition coincides with the convolution form

$$\mathscr{R}_\psi f(\mathbf{u}, \alpha, \beta) = \left(\mathrm{R}f(\mathbf{u}, \cdot) * \overline{\widetilde{\psi_\alpha}} \right)(\beta), \quad (\mathbf{u}, \alpha, \beta) \in \mathbb{Y}^{m+1} \tag{47}$$

where $\psi_\alpha(p) := \psi(p/\alpha)/\alpha$; the convolution $\cdot * \cdot$, dilation $\cdot_\alpha$, reflection $\widetilde{\cdot}$, and complex conjugation $\overline{\cdot}$ are understood as operations for Schwartz distributions.

Theorem 4.2 *(Balancing theorem).* *The ridgelet transform $\mathscr{R} : \mathcal{X}(\mathbb{R}^m) \times \mathcal{Z}(\mathbb{R}) \to \mathcal{Y}(\mathbb{Y}^{m+1})$ is well defined as a bilinear map when $\mathcal{X}$ and $\mathcal{Z}$ are chosen from* Table 4.

The proof is provided in Appendix A. Note that each $\mathcal{Z}$ is (almost) the largest in the sense that the convolution $\mathcal{B} = \mathcal{X} * \mathcal{Z}$ converges. Thus, Table 4 suggests that there is a trade-off relation between $\mathcal{X}$ and $\mathcal{Z}$, that is, as $\mathcal{X}$ increases, $\mathcal{Z}$ decreases and vice versa.

Extension of the ridgelet transform of non-integrable functions requires more sophisticated approaches, because a direct computation of the Radon transform may diverge. For instance, Kostadinova et al. [30] extend $\mathcal{X} = \mathcal{S}_0'$ by using a duality technique. In §5.3 we extend the ridgelet transform to $L^2(\mathbb{R}^m)$, by using the bounded extension procedure.

Proposition 4.3 *(Continuity of the ridgelet transform $L^1(\mathbb{R}^m) \to L^\infty(\mathbb{Y}^{m+1})$). Fix $\psi \in \mathcal{S}(\mathbb{R})$. The ridgelet transform $\mathscr{R}_\psi : L^1(\mathbb{R}^m) \to L^\infty(\mathbb{Y}^{m+1})$ is bounded.*

Proof. Fix an arbitrary $f \in L^1(\mathbb{R}^m)$ and $\psi \in \mathcal{S}(\mathbb{R})$. Recall that this case is absolutely convergent. By using the convolution form,

$$\operatorname*{ess\,sup}_{(\mathbf{u},\alpha,\beta)} \left| \left(\mathrm{R}f(\mathbf{u}, \cdot) * \overline{\widetilde{\psi_\alpha}} \right)(\beta) \right| \leq \|f\|_{L^1(\mathbb{R}^m)} \cdot \operatorname*{ess\,sup}_{(\alpha,\beta)} |\psi_\alpha(\beta)| \tag{48}$$

$$\leq \|f\|_{L^1(\mathbb{R}^m)} \cdot \operatorname*{ess\,sup}_{(r,\beta)} |r \cdot \psi(r\beta)| < \infty, \tag{49}$$

where the first inequality follows by using Young's inequality and applying $\int_\mathbb{R} |\mathrm{R}f(\mathbf{u}, p)| \mathrm{d}p \leq \|f\|_1$; the second inequality follows by changing the variable $r \leftarrow 1/\alpha$, and the resultant is finite because ψ decays rapidly. $\square$

The ridgelet transform $\mathscr{R}_\psi$ is injective when ψ is *admissible*, because if ψ is admissible then the reconstruction formula holds and thus $\mathscr{R}_\psi$ has the inverse. However, $\mathscr{R}_\psi$ is *not* always injective. For instance, take a Laplacian $f := \Delta g$ of some function $g \in \mathcal{S}(\mathbb{R}^m)$ and a polynomial $\psi(z) = z + 1$, which satisfies $\psi^{(2)} \equiv 0$. According to Table 4, $\mathscr{R}_\psi f$ exists as a smooth function because $f \in \mathcal{S}(\mathbb{R}^m)$ and $\psi \in \mathcal{S}'(\mathbb{R})$. In this case $\mathscr{R}_\psi f = 0$, which means $\mathscr{R}_\psi$ is not injective. That is,

$$\mathscr{R}_\psi f(\mathbf{u}, \alpha, \beta) = \left(\mathrm{R}\Delta g(\mathbf{u}, \cdot) * \overline{\widetilde{\psi_\alpha}} \right)(\beta) \tag{50}$$

$$= \left(\partial^2 \mathrm{R}g(\mathbf{u}, \cdot) * \overline{\widetilde{\psi_\alpha}} \right)(\beta) \tag{51}$$

$$= \left(\mathrm{R}g(\mathbf{u}, \cdot) * \partial^2 \overline{\widetilde{\psi_\alpha}} \right)(\beta) \tag{52}$$

$$= \left(\mathrm{R}g(\mathbf{u}, \cdot) * 0 \right)(\beta) \tag{53}$$

$$= 0, \tag{54}$$

where the second equality follows by the intertwining relation $\mathrm{R}\Delta g(\mathbf{u}, p) = \partial_p^2 \mathrm{R}g(\mathbf{u}, p)$ [17]. Clearly the non-injectivity stems from the choice of ψ. In fact, as we see in the next section, no polynomial can be admissible and thus $\mathscr{R}_\psi$ is not injective for any polynomial ψ.

4.2. Dual ridgelet transform with respect to distributions

Definition 4.4 *(Dual ridgelet transform with respect to distributions).* The dual ridgelet transform $\mathscr{R}_\eta^\dagger \mathrm{T}$ of $\mathrm{T} \in \mathcal{Y}(\mathbb{Y}^{m+1})$ with respect to $\eta \in \mathcal{W}(\mathbb{R})$ is given by

$$\mathscr{R}_\eta^\dagger \mathrm{T}(\mathbf{x}) = \lim_{\substack{\delta \to \infty \\ \varepsilon \to 0}} \int_{\mathbb{S}^{m-1}} \int_\varepsilon^\delta \int_\mathbb{R} \mathrm{T}(\mathbf{u}, \alpha, \mathbf{u} \cdot \mathbf{x} - \alpha z) \, \eta(z) \frac{\mathrm{d}z \mathrm{d}\alpha \mathrm{d}\mathbf{u}}{\alpha^m}, \quad \mathbf{x} \in \mathbb{R}^m \tag{55}$$

where $\int_\mathbb{R} \cdot \eta(z) \mathrm{d}z$ is understood as the action of a distribution η.

If the dual ridgelet transform $\mathscr{R}_\eta^\dagger$ exists, then it coincides with the dual operator [36] of the ridgelet transform $\mathscr{R}_\eta$.

Theorem 4.5. *Let $\mathcal{X}$ and $\mathcal{Z}$ be chosen from Table 4. Fix $\psi \in \mathcal{Z}$. Assume that $\mathscr{R}_\psi : \mathcal{X}(\mathbb{R}^m) \to \mathcal{Y}(\mathbb{Y}^{m+1})$ is injective and that $\mathscr{R}_\psi^\dagger : \mathcal{Y}'(\mathbb{Y}^{m+1}) \to \mathcal{X}'(\mathbb{R}^m)$ exists. Then $\mathscr{R}_\psi^\dagger$ is the dual operator $(\mathscr{R}_\psi)' : \mathcal{Y}'(\mathbb{Y}^{m+1}) \to \mathcal{X}'(\mathbb{R}^m)$ of $\mathscr{R}_\psi$.*

Proof. By assumption $\mathscr{R}_\psi$ is densely defined on $\mathcal{X}(\mathbb{R}^m)$ and injective. Therefore, by a classical result on the existence of the dual operator [36, VII. 1. Th. 1, pp.193], there uniquely exists a dual operator $(\mathscr{R}_\psi)' : \mathcal{Y}'(\mathbb{Y}^{m+1}) \to \mathcal{X}'(\mathbb{R}^m)$. On the other hand, for $f \in \mathcal{X}(\mathbb{R}^m)$ and $\mathrm{T} \in \mathcal{Y}(\mathbb{Y}^{m+1})$,

$$\langle \mathscr{R}_\psi f, \mathrm{T} \rangle_{\mathbb{Y}^{m+1}} = \int_{\mathbb{R}^m \times \mathbb{Y}^{m+1}} f(\mathbf{x})\overline{\psi(\mathbf{a} \cdot \mathbf{x} - b)\mathrm{T}(\mathbf{a}, b)}\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{a}\mathrm{d}b = \left\langle f, \mathscr{R}_\psi^\dagger \mathrm{T} \right\rangle_{\mathbb{R}^m}. \tag{56}$$

By the uniqueness of the dual operator, we can conclude $(\mathscr{R}_\psi)' = \mathscr{R}_\psi^\dagger$. $\quad\square$

5. Reconstruction formula for weak ridgelet transform

In this section we discuss the admissibility condition and the reconstruction formula, not only in the Fourier domain as many authors did [23,24,1,30,31], but also in the real domain and in the Radon domain. Both domains are key to the constructive formulation. In §5.1 we derive a constructive admissibility condition. In §5.2 we show two reconstruction formulas. The first of these formulas is obtained by using the Fourier slice theorem and the other by using the Radon transform. In §5.3 we will extend the ridgelet transform to L^2.

5.1. Admissibility condition

Definition 5.1 *(Admissibility condition).* A pair $(\psi, \eta) \in \mathcal{S}(\mathbb{R}) \times \mathcal{S}'(\mathbb{R})$ is said to be admissible when there exists a neighborhood $\Omega \subset \mathbb{R}$ of 0 such that $\widehat{\eta} \in L^1_{\mathrm{loc}}(\Omega \setminus \{0\})$, and the integral

$$K_{\psi,\eta} := (2\pi)^{m-1} \left(\int_{\Omega\setminus\{0\}} + \int_{\mathbb{R}\setminus\Omega} \right) \frac{\overline{\widehat{\psi}(\zeta)}\widehat{\eta}(\zeta)}{|\zeta|^m}\mathrm{d}\zeta, \tag{57}$$

converges and is not zero, where $\int_{\Omega\setminus\{0\}}$ and $\int_{\mathbb{R}\setminus\Omega}$ are understood as Lebesgue's integral and the action of $\widehat{\eta}$, respectively.

Using the Fourier transform in $\mathcal{W}$ requires us to assume that $\mathcal{W} \subset \mathcal{S}'$.

The second integral $\int_{\mathbb{R}\setminus\Omega}$ is always finite because $|\zeta|^{-m} \in \mathcal{O}_M(\mathbb{R}\setminus\Omega)$ and thus $|\zeta|^{-m}\overline{\widehat{\psi}(\zeta)}$ decays rapidly; therefore, by definition the action of a tempered distribution $\widehat{\eta}$ always converges. The convergence of the first integral $\int_{\Omega\setminus\{0\}}$ does not depend on the choice of Ω because for every two neighborhoods Ω and Ω' of 0, the residual $\int_{\Omega\setminus\Omega'}$ is always finite. Hence, the convergence of $K_{\psi,\eta}$ does not depend on the choice of Ω.

The removal of 0 from the integral is essential because a product of two singular distributions, which is indeterminate in general, can occur at 0. See examples below. In Appendix C, we have to treat $|\zeta|^{-m}$ as a locally integrable function, rather than simply a regularized distribution such as Hadamard's finite part. If the integrand coincides with a function at 0, then obviously $\int_{\mathbb{R}\setminus\{0\}} = \int_{\mathbb{R}}$.

If $\widehat{\eta}$ is supported in the singleton $\{0\}$ then η cannot be admissible because $K_{\psi,\eta} = 0$ for any $\psi \in \mathcal{S}(\mathbb{R})$. According to Rudin [34, Ex. 7.16], it happens if and only if η is a polynomial. Therefore, it is natural to take

$\mathcal{W} = \mathcal{S}'/\mathcal{P} \cong \mathcal{S}'_0$ rather than $\mathcal{W} = \mathcal{S}'$. That is, in $\mathcal{S}'_0(\mathbb{R})$, we identify a polynomial $\eta \in \mathcal{P}(\mathbb{R})$ as $0 \in \mathcal{S}'(\mathbb{R})$. The integral $K_{\psi,\eta}$ is well-defined for $\mathcal{S}'_0(\mathbb{R})$. Namely $K_{\psi,\eta}$ is invariant under the addition of a polynomial Q to η

$$K_{\psi,\eta} = K_{\psi,\eta+Q}. \tag{58}$$

Example 5.2 *(Modification of Schwartz [32, Ch. 5, Th. 6]).* Let $\eta(z) = z$ and $\psi(z) = \Lambda \mathrm{G}(z)$ with $\mathrm{G}(z) = \exp(-z^2/2)$. Then,

$$\widehat{\eta}(\zeta) = \delta(\zeta) \quad \text{and} \quad \overline{\widehat{\psi}(\zeta)} = |\zeta| \cdot \mathrm{G}(\zeta). \tag{59}$$

In this case the product of the two distributions is not associative

$$\int_{\mathbb{R}} \mathrm{p.v.} \frac{1}{|\zeta|} \times (|\zeta| \cdot \mathrm{G}(\zeta) \times \delta(\zeta)) \, \mathrm{d}\zeta = 0, \tag{60}$$

$$\int_{\mathbb{R}} \left(\mathrm{p.v.} \frac{1}{|\zeta|} \times |\zeta| \cdot \mathrm{G}(\zeta) \right) \times \delta(\zeta) \mathrm{d}\zeta = \mathrm{G}(0) \neq 0. \tag{61}$$

On the other hand (57) is well defined

$$K_{\psi,\eta} = \int_{0<|\zeta|<1} \frac{|\zeta| \cdot \mathrm{G}(\zeta) \times 0}{|\zeta|} \mathrm{d}\zeta + \int_{1 \le |\zeta|} \frac{|\zeta| \cdot \mathrm{G}(\zeta)}{|\zeta|} \delta(\zeta) \mathrm{d}\zeta = 0. \tag{62}$$

Example 5.3. Let $\eta(z) = z_+^0 + (2\pi)^{-1} \exp iz$ and $\psi(z) = \Lambda \mathrm{G}(z)$. Then,

$$\widehat{\eta}(\zeta) = \frac{i}{\zeta} + \delta(\zeta) + \delta(\zeta - 1) \quad \text{and} \quad \overline{\widehat{\psi}(\zeta)} = |\zeta| \cdot \mathrm{G}(\zeta). \tag{63}$$

The product of the two distributions is not associative

$$\int_{\mathbb{R}} \mathrm{p.v.} \frac{1}{|\zeta|} \times \left(|\zeta| \cdot \mathrm{G}(\zeta) \times \left(\frac{i}{\zeta} + \delta(\zeta) + \delta(\zeta - 1) \right) \right) \mathrm{d}\zeta = \mathrm{G}(1), \tag{64}$$

$$\int_{\mathbb{R}} \left(\mathrm{p.v.} \frac{1}{|\zeta|} \times |\zeta| \cdot \mathrm{G}(\zeta) \right) \times \left(\frac{i}{\zeta} + \delta(\zeta) + \delta(\zeta - 1) \right) \mathrm{d}\zeta = \mathrm{G}(0) + \mathrm{G}(1) \neq 0. \tag{65}$$

On the other hand, (57) is well defined

$$K_{\psi,\eta} = \int_{0<|\zeta|<1} \frac{|\zeta| \cdot \mathrm{G}(\zeta) \times i\zeta^{-1}}{|\zeta|} \mathrm{d}\zeta + \int_{1 \le |\zeta|} \frac{|\zeta| \cdot \mathrm{G}(\zeta)}{|\zeta|} \left(\mathrm{p.v.} \frac{i}{\zeta} + \delta(\zeta) + \delta(\zeta - 1) \right) \mathrm{d}\zeta = \infty + \mathrm{G}(1). \tag{66}$$

Observe that formally the integrand $\widehat{u}(\zeta) := \overline{\widehat{\psi}(\zeta)} \widehat{\eta}(\zeta) |\zeta|^{-m}$ is a solution of $|\zeta|^m \widehat{u}(\zeta) = \overline{\widehat{\psi}(\zeta)} \widehat{\eta}(\zeta)$. By taking the Fourier inversion, we have $\Lambda^m u = \overline{\widetilde{\psi}} * \eta$. To be exact, $\widehat{\eta}$ may contain a point mass at the origin, such as Dirac's δ.

Theorem 5.4 *(Structure theorem for admissible pairs).* Let $(\psi, \eta) \in \mathcal{S}(\mathbb{R}) \times \mathcal{S}'(\mathbb{R})$. Assume that there exists $k \in \mathbb{N}_0$ such that

$$\widehat{\eta}(\zeta) = \sum_{j=0}^{k} c_j \delta^{(j)}(\zeta), \quad \zeta \in \{0\}. \tag{67}$$

Assume that there exists a neighborhood Ω of 0 such that $\widehat{\eta} \in C^0(\Omega \setminus \{0\})$. Then ψ and η are admissible if and only if there exists $u \in \mathcal{O}_{\mathcal{M}}(\mathbb{R})$ such that

$$\Lambda^m u = \overline{\widetilde{\psi}} * \left(\eta - \sum_{j=0}^{k} c_j z^j \right) \quad \text{and} \quad \int_{\mathbb{R} \setminus \{0\}} \widehat{u}(\zeta) \mathrm{d}\zeta \neq 0, \tag{68}$$

where Λ is the backprojection filter defined in (24). *In addition,* $\lim_{\zeta \to +0} |\widehat{u}(\zeta)| < \infty$ *and* $\lim_{\zeta \to -0} |\widehat{u}(\zeta)| < \infty$.

The proof is provided in Appendix B. Note that the continuity implies local integrability. If ψ has ℓ vanishing moments with $\ell \geq k$, namely $\int_{\mathbb{R}} \psi(z) z^j \mathrm{d}z = 0$ for $j \leq \ell$, then the condition reduces to

$$\Lambda^m u = \overline{\widetilde{\psi}} * \eta, \quad \left| \int_{\mathbb{R}} u(z) \mathrm{d}z \right| < \infty \quad \text{and} \quad \int_{\mathbb{R}} \widehat{u}(\zeta) \mathrm{d}\zeta \neq 0. \tag{69}$$

As a consequence of Theorem 5.4, we can construct admissible pairs as below.

Corollary 5.5 *(Construction of admissible pairs). Given $\eta \in \mathcal{S}'_0(\mathbb{R})$. Assume that there exists a neighborhood Ω of 0 and $k \in \mathbb{N}_0$ such that $\zeta^k \cdot \widehat{\eta}(\zeta) \in C^0(\Omega)$. Take $\psi_0 \in \mathcal{S}(\mathbb{R})$ such that*

$$\int_{\mathbb{R}} \zeta^k \overline{\widehat{\psi_0}(\zeta)} \widehat{\eta}(\zeta) \mathrm{d}\zeta \neq 0. \tag{70}$$

Then

$$\psi := \Lambda^m \psi_0^{(k)}, \tag{71}$$

is admissible with η.

The proof is obvious because $u := \overline{\widetilde{\psi_0^{(k)}}} * \eta$ satisfies the conditions in Theorem 5.4.

5.2. Reconstruction formula

Theorem 5.6 *(Reconstruction formula). Let $f \in L^1(\mathbb{R}^m)$ satisfy $\widehat{f} \in L^1(\mathbb{R}^m)$ and let $(\psi, \eta) \in \mathcal{S}(\mathbb{R}) \times \mathcal{S}'_0(\mathbb{R})$ be admissible. Then the reconstruction formula*

$$\mathscr{R}_\eta^\dagger \mathscr{R}_\psi f(\mathbf{x}) = K_{\psi,\eta} f(\mathbf{x}), \tag{72}$$

holds for almost every $\mathbf{x} \in \mathbb{R}^m$. The equality holds for every point where f is continuous.

The proof is provided in Appendix C. The admissibility condition can be easily inverted to $(\psi, \eta) \in \mathcal{S}'_0 \times \mathcal{S}$. However, extensions to $\mathcal{S}'_0 \times \mathcal{S}'_0$ and $\mathcal{S} \times \mathcal{D}'$ may *not* be easy. This is because the multiplication $\mathcal{S}'_0 \cdot \mathcal{S}'_0$ is not always commutative, nor associative, and the Fourier transform is not always defined over $\mathcal{D}'$ [32].

The following theorem is another suggestive reconstruction formula that implies wavelet analysis in the Radon domain works as a backprojection filter. In other words, the admissibility condition requires (ψ, η) to construct the filter Λ^m. Note that similar techniques are obtained for "wavelet measures" by Rubin [29,25].

Theorem 5.7 (*Reconstruction formula via radon transform*). *Let $f \in L^1(\mathbb{R}^m)$ be sufficiently smooth and $(\psi, \eta) \in \mathcal{S}(\mathbb{R}) \times \mathcal{S}'(\mathbb{R})$ be admissible. Assume that there exists a real-valued smooth and integrable function u such that*

$$\Lambda^m u = \overline{\overline{\psi}} * \eta \quad and \quad \int_{\mathbb{R}} \widehat{u}(\zeta) \mathrm{d}\zeta = -1. \tag{73}$$

Then,

$$\mathscr{R}_\eta^\dagger \mathscr{R}_\psi f(\mathbf{x}) = \mathrm{R}^* \Lambda^{m-1} \mathrm{R} f(\mathbf{x}) = 2(2\pi)^{m-1} f(\mathbf{x}), \tag{74}$$

holds for almost every $\mathbf{x} \in \mathbb{R}^m$.

The proof is provided in Appendix D. Note that here we imposed a stronger condition on u than the $u \in L^1(\mathbb{R} \setminus \{0\})$ we imposed in Theorem 5.4.

Recall intertwining relations ([17, Lem. 2.1, Th. 3.1, Th. 3.7])

$$(-\Delta)^{\frac{m-1}{2}} \mathrm{R}^* = \Lambda^{m-1} \mathrm{R}, \quad and \quad \mathrm{R}(-\Delta)^{\frac{m-1}{2}} = \mathrm{R}^* \Lambda^{m-1}. \tag{75}$$

Therefore, we have the following.

Corollary 5.8.

$$\mathscr{R}_\eta^\dagger \mathscr{R}_\psi = \mathrm{R}^* \Lambda^{m-1} \mathrm{R} = (-\Delta)^{\frac{m-1}{2}} \mathrm{R}^* \mathrm{R} = \mathrm{R}^* \mathrm{R}(-\Delta)^{\frac{m-1}{2}}. \tag{76}$$

5.3. Extension to L^2

By $(\cdot, \cdot)$ and $\| \cdot \|_2$, with a slight abuse of notation, we denote the inner product of $L^2(\mathbb{R}^m)$ and $L^2(\mathbb{Y}^{m+1})$. Here we endow $\mathbb{Y}^{m+1}$ with a fixed measure $\alpha^{-m} \mathrm{d}\alpha \mathrm{d}\beta \mathrm{d}\mathbf{u}$, and omit writing it explicitly as $L^2(\mathbb{Y}^{m+1}; \dots)$. We say that ψ is *self-admissible* if ψ is admissible in itself, i.e. the pair (ψ, ψ) is admissible. The following relation is immediate by the duality.

Theorem 5.9 (*Parseval's relation and Plancherel's identity*). *Let $(\psi, \eta) \in \mathcal{S} \times \mathcal{S}'$ be admissible with, for simplicity, $K_{\psi, \eta} = 1$. For $f, g \in L^1 \cap L^2(\mathbb{R}^m)$,*

$$(\mathscr{R}_\psi f, \mathscr{R}_\eta g) = \left(\mathscr{R}_\eta^\dagger \mathscr{R}_\psi f, g\right) = (f, g). \qquad Parseval's\ Relation \tag{77}$$

In particular, if ψ is self-admissible, then

$$\|\mathscr{R}_\psi f\|_2 = \|f\|_2. \qquad Plancherel's\ identity \tag{78}$$

Recall Proposition 4.3 that the ridgelet transform is a bounded linear operator on $L^1(\mathbb{R}^m)$. If $\psi \in \mathcal{S}(\mathbb{R})$ is self-admissible, then we can extend the ridgelet transform to $L^2(\mathbb{R}^m)$, by following the bounded extension procedure [40, 2.2.4]. That is, for $f \in L^2(\mathbb{R}^m)$, take a sequence $f_n \in L^1 \cap L^2(\mathbb{R}^m)$ such that $f_n \to f$ in L^2. Then by Plancherel's identity,

$$\|f_n - f_m\|_2 = \|\mathscr{R}_\psi f_n - \mathscr{R}_\psi f_m\|_2, \quad \forall n, m \in \mathbb{N}. \tag{79}$$

The right-hand side is a Cauchy sequence in $L^2(\mathbb{Y}^{m+1})$ as $n, m \to \infty$. By the completeness, there uniquely exists the limit $\mathrm{T}_\infty \in L^2(\mathbb{Y}^{m+1})$ of $\mathscr{R}_\psi f_n$. We regard T_∞ as the ridgelet transform of f and define $\mathscr{R}_\psi f :=$ T_∞.

Theorem 5.10 (*Bounded extension of ridgelet transform on L^2*). *Let $\psi \in \mathcal{S}(\mathbb{R})$ be self-admissible with $K_{\psi,\psi=1}$. The ridgelet transform on $L^1 \cap L^2(\mathbb{R}^m)$ admits a unique bounded extension to $L^2(\mathbb{R}^m)$, with satisfying $\|\mathscr{R}_\psi f\|_2 = \|f\|_2$.*

We say that (ψ, η) and $(\psi^\star, \eta^\star)$ are *equivalent*, if two admissible pairs (ψ, η) and $(\psi^\star, \eta^\star)$ define the same convolution $\overline{\tilde{\psi}} * \eta = \overline{\tilde{\psi^\star}} * \eta^\star$ in common. If (ψ, η) and $(\psi^\star, \eta^\star)$ are equivalent, then obviously

$$(\mathscr{R}_\psi f, \mathscr{R}_\eta g) = (\mathscr{R}_{\psi^\star} f, \mathscr{R}_{\eta^\star} g). \tag{80}$$

We say that an admissible pair (ψ, η) is *admissibly decomposable*, when there exist self-admissible pairs $(\psi^\star, \psi^\star)$ and $(\eta^\star, \eta^\star)$ such that $(\psi^\star, \eta^\star)$ is equivalent to (ψ, η). If (ψ, η) is admissibly decomposable with $(\psi^\star, \eta^\star)$, then by the Schwartz inequality

$$(\mathscr{R}_\psi f, \mathscr{R}_\eta g) \leq \|\mathscr{R}_{\psi^\star} f\|_2 \|\mathscr{R}_{\eta^\star} g\|_2. \tag{81}$$

Theorem 5.11 (*Reconstruction formula in L^2*). *Let $f \in L^2(\mathbb{R}^m)$ and $(\psi, \eta) \in \mathcal{S} \times \mathcal{S}'$ be admissibly decomposable with $K_{\psi,\eta} = 1$. Then,*

$$\mathscr{R}_\eta^\dagger \mathscr{R}_\psi f \to f, \quad \text{in } L^2. \tag{82}$$

The proof is provided in Appendix E. Even when ψ is not self-admissible and thus $\mathscr{R}_\psi$ cannot be defined on $L^2(\mathbb{R}^m)$, the reconstruction operator $\mathscr{R}_\eta^\dagger \mathscr{R}_\psi$ can be defined with the aid of η.

6. Neural network with unbounded activation functions

In this section we instantiate the universal approximation property for the variants of neural networks. Recall that a neural network coincides with the dual ridgelet transform of a function. Henceforth, we rephrase a dual ridgelet function as an activation function. According to the reconstruction formulas (Theorem 5.6, 5.7, and 5.11), we can determine whether a neural network with an activation function η is a universal approximator by checking the admissibility of η.

Table 1 lists some Lizorkin distributions for potential activation functions. In §6.1 we verify that they belong to $\mathcal{S}_0'(\mathbb{R})$ and some of them belong to $\mathcal{O}_M(\mathbb{R})$ and $\mathcal{S}(\mathbb{R})$, which are subspaces of $\mathcal{S}_0'(\mathbb{R})$. In §6.2 we show that they are admissible with some ridgelet function $\psi \in \mathcal{S}(\mathbb{R})$; therefore, each of their corresponding neural networks is a universal approximator.

6.1. Examples of Lizorkin distributions

We proved the class properties by using the following propositions.

Proposition 6.1 (*Tempered distribution $\mathcal{S}'(\mathbb{R})$ [40, Ex. 2.3.5]*). *Let $g \in L^1_{\text{loc}}(\mathbb{R})$. If $|g(z)| \lesssim (1 + |z|)^k$ for some $k \in \mathbb{N}_0$, then $g \in \mathcal{S}'(\mathbb{R})$.*

Proposition 6.2 (*Slowly increasing function $\mathcal{O}_M(\mathbb{R})$ [40, Def. 2.3.15]*). *Let $g \in \mathcal{E}(\mathbb{R})$. If for any $\alpha \in \mathbb{N}_0$, $|\partial^\alpha g(x)| \lesssim (1 + |z|)^{k_\alpha}$ for some $k_\alpha \in \mathbb{N}_0$, then $g \in \mathcal{O}_M(\mathbb{R})$.*

Example 6.3. Truncated power functions z_+^k $(k \in \mathbb{N}_0)$, which contain the ReLU z_+ and the step function z_+^0, belong to $\mathcal{S}_0'(\mathbb{R})$.

Proof. For any $\ell \in \mathbb{N}_0$ there exists a constant C_ℓ such that $|\partial^\ell (z_+^k)| \leq C_\ell (1 + |z|)^{k-\ell}$. Hence, $z_+^k \in \mathcal{S}_0'(\mathbb{R})$. $\quad\square$

Example 6.4. The sigmoidal function $\sigma(z)$ and the softplus $\sigma^{(-1)}(z)$ belong to $\mathcal{O}_{\mathcal{M}}(\mathbb{R})$. The derivatives $\sigma^{(k)}(z)$ $(k \in \mathbb{N})$ belong to $\mathcal{S}(\mathbb{R})$. Hyperbolic tangent $\tanh(z)$ belongs to $\mathcal{O}_{\mathcal{M}}(\mathbb{R})$.

The proof is provided in Appendix F.

Example 6.5. (See [40, Ex. 2.2.2].) RBF $\mathrm{G}(z)$ and their derivatives $\mathrm{G}^{(k)}(z)$ belong to $\mathcal{S}(\mathbb{R})$.

Example 6.6. (See [40, Ex. 2.3.5].) Dirac's $\delta(z)$ and their derivatives $\delta^{(k)}(z)$ belong to $\mathcal{S}'(\mathbb{R})$.

6.2. $K_{\psi,\eta}$ when ψ is a derivative of the Gaussian

Given an activation function $\eta \in \mathcal{S}'_0(\mathbb{R})$, according to Corollary 5.5 we can construct an admissible ridgelet function $\psi \in \mathcal{S}(\mathbb{R})$ by letting

$$\psi := \Lambda^m \psi_0, \tag{83}$$

where $\psi_0 \in \mathcal{S}(\mathbb{R})$ satisfies

$$\left\langle \widehat{\eta}, \widehat{\psi_0} \right\rangle := \int_{\mathbb{R}\backslash\{0\}} \overline{\widehat{\psi_0}(\zeta)}\widehat{\eta}(\zeta)\mathrm{d}\zeta \neq 0, \pm\infty. \tag{84}$$

Here we consider the case when ψ_0 is given by

$$\psi_0 = \mathrm{G}^{(\ell)}, \tag{85}$$

for some $\ell \in \mathbb{N}_0$, where G denotes the Gaussian $\mathrm{G}(z) := \exp(-z^2/2)$.

The Fourier transform of the Gaussian is given by $\widehat{\mathrm{G}}(\zeta) = \exp(-\zeta^2/2) = \sqrt{2\pi}\,\mathrm{G}(\zeta)$. The Hilbert transform of the Gaussian, which we encounter by computing $\psi = \Lambda^m \mathrm{G}$ when m is odd, is given by

$$\mathcal{H}\mathrm{G}(z) = \frac{2i}{\sqrt{\pi}}F\left(\frac{z}{\sqrt{2}}\right), \tag{86}$$

where $F(z)$ is the Dawson function $F(z) := \exp(-z^2)\int_0^z \exp(w^2)\mathrm{d}w$.

Example 6.7. z_+^k $(k \in \mathbb{N}_0)$ is admissible with $\psi = \Lambda^m \mathrm{G}^{(\ell+k+1)}$ $(\ell \in \mathbb{N}_0)$ iff ℓ is even. If odd, then $K_{\psi,\eta} = 0$.

Proof. It follows from the fact that, according to Gel'fand and Shilov [41, § 9.3],

$$\widehat{z_+^k}(\zeta) = \frac{k!}{(i\zeta)^{k+1}} + \pi i^k \delta^{(k)}(\zeta), \quad k \in \mathbb{N}_0. \qquad \square \tag{87}$$

Example 6.8. $\eta(z) = \delta^{(k)}(z)$ $(k \in \mathbb{N}_0)$ is admissible with $\psi = \Lambda^m \mathrm{G}$ iff k is even. If odd, then $K_{\psi,\eta} = 0$.

In contrast to polynomial functions, Dirac's δ can be an admissible activation function.

Example 6.9. $\eta(z) = \mathrm{G}^{(k)}(z)$ $(k \in \mathbb{N}_0)$ is admissible with $\psi = \Lambda^m \mathrm{G}$ iff k is even. If odd, then $K_{\psi,\eta} = 0$.

Example 6.10. $\eta(z) = \sigma^{(k)}(z)$ $(k \in \mathbb{N}_0)$ is admissible with $\psi = \Lambda^m \mathrm{G}$ iff k is odd. If odd, then $K_{\psi,\eta} = 0$. $\sigma^{(-1)}$ is admissible with $\psi = \Lambda^m \mathrm{G}''$.

The proof is provided in Appendix F.

Table 5
Theoretical diagnoses for admissibility of $\psi = \Lambda^m \psi_0$ and η. '+' indicates that (ψ, η) is admissible. '0' and '∞' indicate that $K_{\psi,\eta}$ vanishes and diverges, respectively, and thus (ψ, η) is not admissible.

Activation function	η	$\psi = \Lambda^m \mathrm{G}$	$\psi = \Lambda^m \mathrm{G}'$	$\psi = \Lambda^m \mathrm{G}''$
Derivative of sigmoidal ft.	σ'	+	0	+
Sigmoidal function	σ	∞	+	0
Softplus	$\sigma^{(-1)}$	∞	∞	+
Dirac's δ	δ	+	0	+
Unit step function	z_+^0	∞	+	0
ReLU	z_+	∞	∞	+
Linear function	z	0	0	0
RBF	G	+	0	+

7. Numerical examples of reconstruction

We performed some numerical experiments on reconstructing a one-dimensional signal and a two-dimensional image, with reference to our theoretical diagnoses for admissibility in the previous section. Table 5 lists the diagnoses of $(\Lambda^m \psi_0, \eta)$ we employ in this section. The symbols '+,' '0,' and '∞' in each cell indicate that $K_{\psi,\eta}$ of the corresponding (ψ, η) converges to a non-zero constant (+), converges to zero (0), and diverges (∞). Hence, by Theorem 5.6, if the cell (ψ, η) indicates '+' then a neural network with an activation function η is a universal approximator.

7.1. Sinusoidal curve

We studied a one-dimensional signal $f(x) = \sin 2\pi x$ defined on $x \in [-1, 1]$. The ridgelet functions $\psi = \Lambda \psi_0$ were chosen from derivatives of the Gaussian $\psi_0 = \mathrm{G}^{(\ell)}$, $(\ell = 0, 1, 2)$. The activation functions η were chosen from among the softplus $\sigma^{(-1)}$, the sigmoidal function σ and its derivative σ', the ReLU z_+, unit step function z_+^0, and Dirac's δ. In addition, we examined the case when the activation function is simply a linear function: $\eta(z) = z$, which cannot be admissible because the Fourier transform of polynomials is supported at the origin in the Fourier domain.

The signal was sampled from $[-1, 1]$ with $\Delta x = 1/100$. We computed the reconstruction formula

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \mathscr{R}_\psi f(a,b) \eta(a \cdot x - b) \frac{\mathrm{d}a\mathrm{d}b}{|a|}, \tag{88}$$

by simply discretizing $(a, b) \in [-30, 30] \times [-30, 30]$ by $\Delta a = \Delta b = 1/10$. That is,

$$\mathscr{R}_\psi f(a,b) \approx \sum_{n=0}^{N} f(x_n) \overline{\psi(a \cdot x_n - b)} |a| \Delta x, \quad x_n = x_0 + n\Delta x \tag{89}$$

$$\mathscr{R}_\eta^\dagger \mathscr{R}_\psi f(x) \approx \sum_{(i,j)=(0,0)}^{I,J} \mathscr{R}_\psi f(a_i, b_j) \eta(a_i \cdot x - b_j) \frac{\Delta a \Delta b}{|a_i|}, \quad a_i = a_0 + i\Delta a, \ b_j = b_0 + j\Delta b \tag{90}$$

where $x_0 = -1$, $a_0 = -30$, $b_0 = -30$, and $N = 200, (I, J) = (600, 600)$.

Fig. 2 depicts the ridgelet transform $\mathscr{R}_\psi f(a, b)$. As the order ℓ of $\psi = \Lambda \mathrm{G}^{(\ell)}$ increases, the localization of $\mathscr{R}_\psi f$ increases. As shown in Fig. 3, every $\mathscr{R}_\psi f$ can be reconstructed to f with some admissible activation function η. It is somewhat intriguing that the case $\psi = \Lambda \mathrm{G}''$ can be reconstructed with two different activation functions.

Figs. 3, 4, and 5 tile the results of reconstruction with sigmoidal functions, truncated power functions, and a linear function. The solid line is a plot of the reconstruction result; the dotted line draws the original
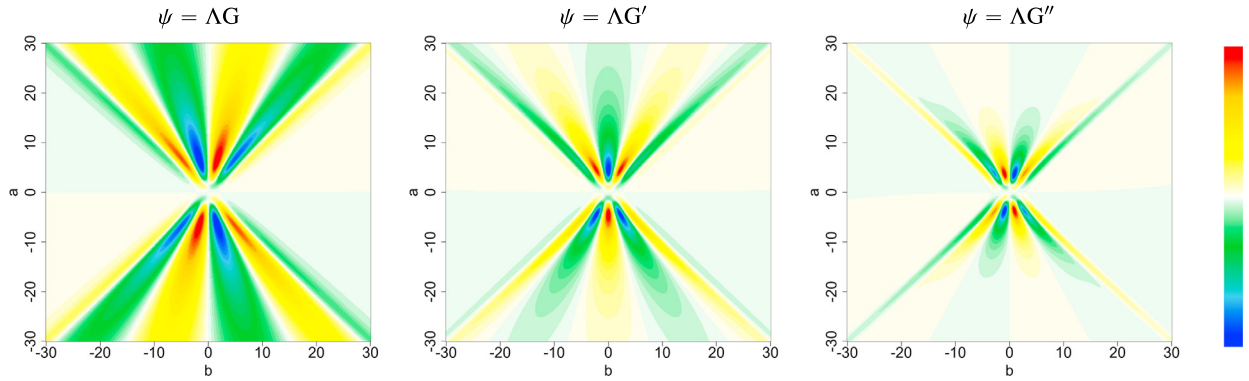
Fig. 2. Ridgelet transform $\mathscr{R}_\psi f(a,b)$ of $f(x) = \sin 2\pi x$ defined on $[-1,1]$ with respect to ψ.
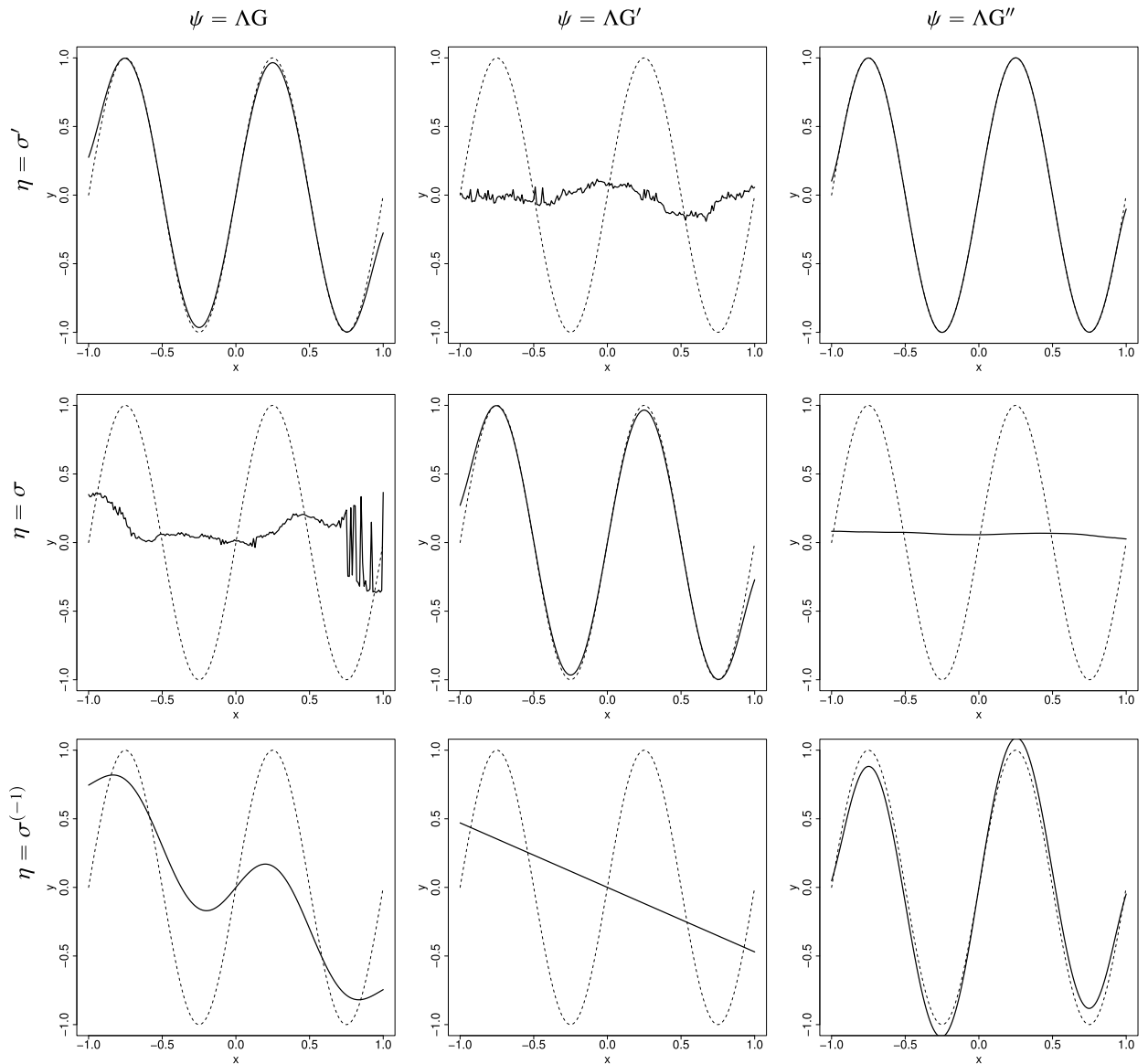

Fig. 3. Reconstruction with the derivative of sigmoidal function σ', sigmoidal function σ, and softplus $\sigma^{(-1)}$. The solid line plots the reconstruction result; the dotted line plots the original signal.
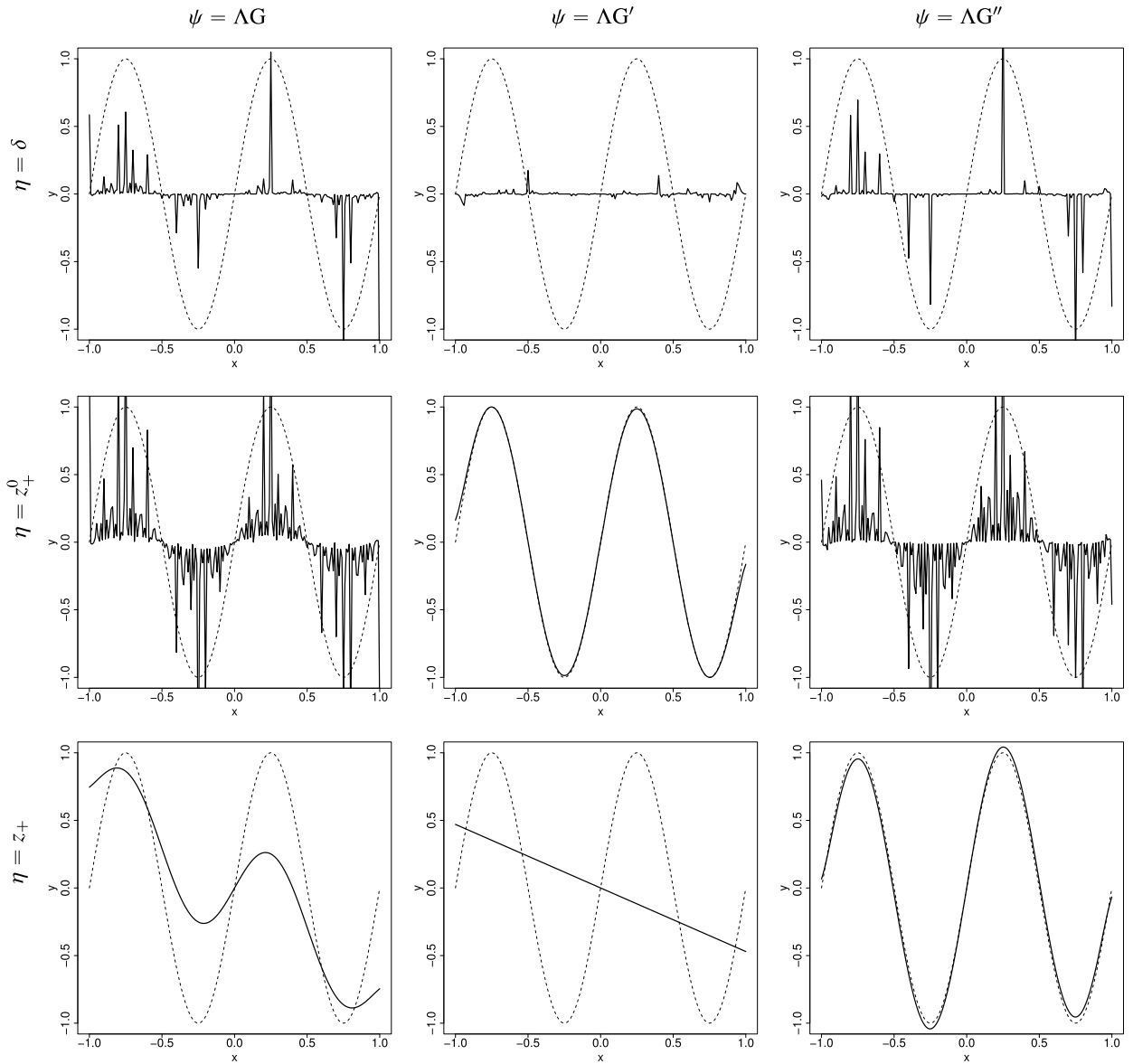
$\psi = \Lambda G$ $\psi = \Lambda G'$ $\psi = \Lambda G''$

Fig. 4. Reconstruction with truncated power functions — Dirac's δ, unit step z_+^0, and ReLU z_+. The solid line plots the reconstruction result; the dotted line plots the original signal.
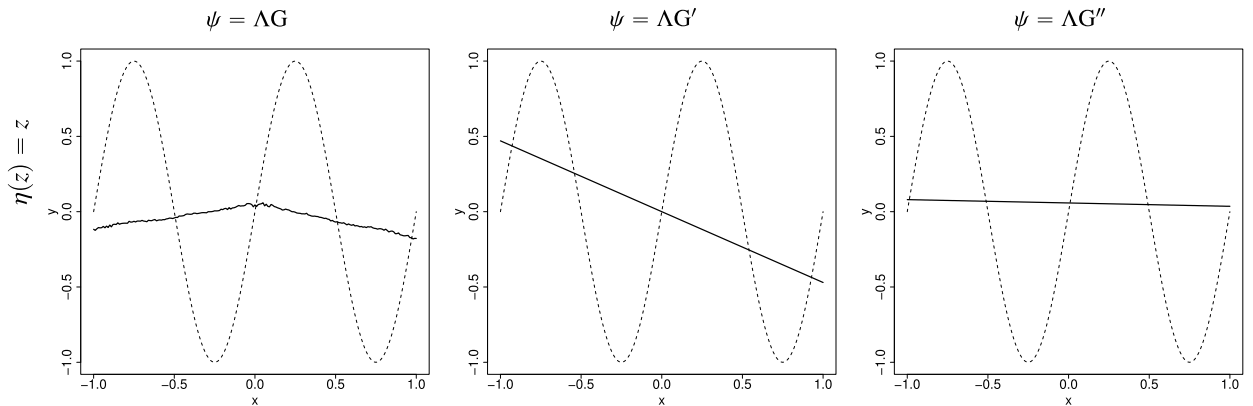
Fig. 5. Reconstruction with linear function $\eta(z) = z$. The solid line plots the reconstruction result; the dotted line plots the original signal.

signal. In each of the figures, the theoretical diagnoses and experimental results are almost consistent and reasonable.

In Fig. 3, at the bottom left, the reconstruction signal with the softplus seems incompletely reconstructed, in spite of Table 5 indicating '∞'. Recall that $\widehat{\sigma^{(-1)}}(\zeta)$ has a pole ζ^{-2}; thus, we can understand this cell in terms of $\sigma^{(-1)} * \Lambda G$ working as an *integrator*, that is, a *low-pass filter*.

In Fig. 4, in the top row, all the reconstructions with Dirac's δ fail. These results seem to contradict the theory. However, it simply reflects the implementation difficulty of realizing Dirac's δ, because $\delta(z)$ is a "function" that is almost constantly zero, except for the origin. Nevertheless, $z = ax - b$ rarely happens to be exactly zero, provided a, b, and x are discretized. This is the reason why this row fails. At the bottom left, the ReLU seems to lack sharpness for reconstruction. Here we can again understand that $z_+ * \Lambda G$ worked as a low-pass filter. It is worth noting that the unit step function and the ReLU provide a sharper reconstruction than the sigmoidal function and the softplus.

In Fig. 5, all the reconstructions with a linear function fail. This is consistent with the theory that polynomials cannot be admissible as their Fourier transforms are singular at the origin.

7.2. Shepp–Logan phantom

We next studied a gray-scale image *Shepp–Logan phantom* [42]. The ridgelet functions $\psi = \Lambda^2 \psi_0$ were chosen from the ℓth derivatives of the Gaussian $\psi_0 = G^{(\ell)}$, $(\ell = 0, 1, 2)$. The activation functions η were chosen from the RBF G (instead of Dirac's δ), the unit step function z_+^0, and the ReLU z_+.

The original image was composed of 256×256 pixels. We treated it as a two-dimensional signal $f(\mathbf{x})$ defined on $[-1, 1]^2$. We computed the reconstruction formula

$$\int_{\mathbb{R}} \int_{\mathbb{R}^2} \mathscr{R}_\psi f(\mathbf{a}, b) \eta(\mathbf{a} \cdot \mathbf{x} - b) \frac{\mathrm{d}\mathbf{a}\mathrm{d}b}{\|\mathbf{a}\|}, \tag{91}$$

by discretizing $(\mathbf{a}, b) \in [-300, 300]^2 \times [-30, 30]$ by $\Delta\mathbf{a} = (1, 1)$ and $\Delta b = 1$.

Fig. 6 lists the results of the reconstruction. As observed in the one-dimensional case, the results are fairly consistent with the theory. Again, at the bottom left, the reconstructed image seems dim. Our understanding is that it was caused by low-pass filtering.

8. Concluding remarks

We have shown that neural networks with unbounded non-polynomial activation functions have the universal approximation property. Because the integral representation of the neural network coincides with the dual ridgelet transform, our goal reduces to constructing the ridgelet transform with respect to distributions. Our results cover a wide range of activation functions: not only the traditional RBF, sigmoidal function, and unit step function, but also truncated power functions z_+^k, which contain the ReLU and even Dirac's δ. In particular, we concluded that a neural network can approximate $L^1 \cap C^0$ functions in the pointwise sense, and L^2 functions in the L^2 sense, when its activation "function" is a Lizorkin distribution $(\mathcal{S}_0')$ that is admissible. The Lizorkin distribution is a tempered distribution $(\mathcal{S}')$ that is not a polynomial. As an important consequence, what a neural network learns is a ridgelet transform of the target function f. In other words, during backpropagation the network indirectly searches for an admissible ridgelet function, by constructing a backprojection filter.

Using the weak form expression of the ridgelet transform, we extensively defined the ridgelet transform with respect to distributions. Theorem 4.2 guarantees the existence of the ridgelet transform with respect to distributions. Table 4 suggests that for the convolution of distributions to converge, the class $\mathcal{X}$ of domain and the class $\mathcal{Z}$ of ridgelets should be balanced. Proposition 4.3 states that $\mathscr{R}_\psi : L^1(\mathbb{R}^m) \to \mathcal{S}'(\mathbb{Y}^{m+1})$
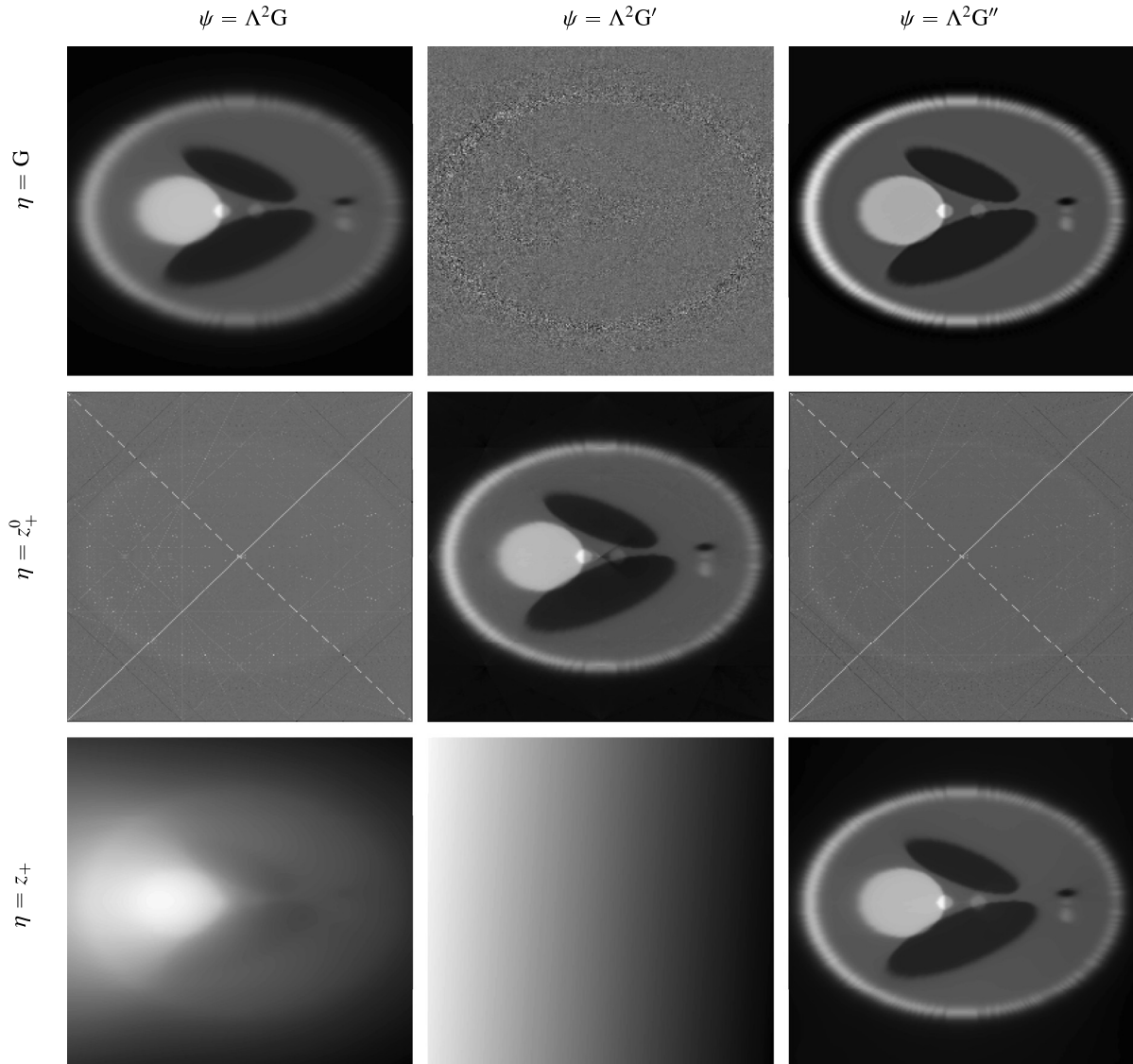
Fig. 6. Reconstruction with RBF G, unit step z_+^0, and ReLU z_+.

is a bounded linear operator. Theorem 4.5 states that the dual ridgelet transform coincides with a dual operator. Provided the reconstruction formula holds, that is, when the ridgelets are admissible, the ridgelet transform is injective and the dual ridgelet transform is surjective.

For an unbounded $\eta \in \mathcal{Z}(\mathbb{R})$ to be admissible, it cannot be a polynomial and it can be associated with a backprojection filter. If $\eta \in \mathcal{Z}(\mathbb{R})$ is a polynomial then the product of distributions in the admissibility condition should be indeterminate. Therefore, $\mathcal{Z}(\mathbb{R})$ excludes polynomials. Theorem 5.4 rephrases the admissibility condition in the real domain. As a direct consequence, Corollary 5.5 gives a constructive sufficiently admissible condition.

After investigating the construction of the admissibility condition, we showed that formulas can be reconstructed on $L^1(\mathbb{R}^m)$ in two ways. Theorem 5.6 uses the Fourier slice theorem. Theorem 5.7 uses approximations to the identity and reduces to the inversion formula of the Radon transform. Theorem 5.7 as well as Corollary 5.8 suggest that the admissibility condition requires (ψ, η) to construct a backprojection filter.

In addition, we have extended the ridgelet transform on $L^1(\mathbb{R}^m)$ to $L^2(\mathbb{R}^m)$. Theorem 5.9 states that Parseval's relation, which is a weak version of the reconstruction formula, holds on $L^1 \cap L^2(\mathbb{R}^m)$. Theorem 5.10

follows the bounded extension of $\mathscr{R}_\psi$ from $L^1 \cap L^2(\mathbb{R}^m)$ to $L^2(\mathbb{R}^m)$. Theorem 5.11 gives the reconstruction formula in $L^2(\mathbb{R}^m)$.

By showing that z_+^k and other activation functions belong to $\mathcal{S}_0'$, and that they are admissible with some derivatives of the Gaussian, we proved the universal approximation property of a neural network with an unbounded activation function. Numerical examples were consistent with our theoretical diagnoses on the admissibility. In addition, we found that some non-admissible combinations worked as a low-pass filter; for example, $(\psi, \eta) = (\Lambda^m[\text{Gaussian}], \text{ReLU})$ and $(\psi, \eta) = (\Lambda^m[\text{Gaussian}], \text{softplus})$.

We plan to perform the following interesting investigations in future.

1. Given an activation function $\eta \in \mathcal{S}_0'(\mathbb{R})$, which is the "best" ridgelet function $\psi \in \mathcal{S}(\mathbb{R})$?

    In fact, for a given activation function η, we have plenty of choices. By Corollary 5.5, all elements of

$$\mathcal{A}_\eta := \left\{ \Lambda^m \psi_0 \;\middle|\; \psi_0 \in \mathcal{S}(\mathbb{R}) \text{ such that } \langle \widehat{\eta}, \widehat{\psi_0} \rangle \text{ is finite and nonzero.} \right\}, \tag{92}$$

    are admissible with η.

2. How are ridgelet functions related to *deep* neural networks?

    Because ridgelet analysis is so fruitful, we aim to develop "deep" ridgelet analysis. One of the essential leaps from shallow to deep is that the network output expands from scalar to vector because a deep structure is a cascade of multi-input multi-output layers. In this regard, we expect Corollary 5.8 to play a key role. By using the intertwining relations, we can "cascade" the reconstruction operators as below

$$\mathscr{R}_\eta^\dagger \mathscr{R}_\psi \mathscr{R}_\eta^\dagger \mathscr{R}_\psi = \mathrm{R}^* \Lambda^{k-1} \mathrm{R}(-\Delta)^{m-1-\frac{k+\ell}{2}} \mathrm{R}^* \Lambda^{\ell-1} \mathrm{R}. \quad (0 \le k, \ell \le m) \tag{93}$$

    This equation suggests that the cascade of ridgelet transforms coincides with a composite of back-projection filtering in the Radon domain and differentiation in the real domain. We conjecture that this point of view can be expected to facilitate analysis of the deep structure.

Acknowledgments

Appendix A. Proof of Theorem 4.2

A ridgelet transform $\mathscr{R}_\psi f(\mathbf{u}, \alpha, \beta)$ is the convolution of a Radon transform $\mathrm{R}f(\mathbf{u}, p)$ and a dilated distribution $\psi_\alpha(p)$ in the sense of a Schwartz distribution. That is,

$$f(\mathbf{x}) \mapsto \mathrm{R}f(\mathbf{u}, p) \mapsto \left(\mathrm{R}f(\mathbf{u}, \cdot) * \overline{\overline{\psi_\alpha}} \right)(\beta) = \mathscr{R}_\psi f(\mathbf{u}, \alpha, \beta). \tag{A.1}$$

We verify that the ridgelet transform is well defined in a stepwise manner. Provided there is no danger of confusion, in the following steps we denote by $\mathcal{X}$ the classes $\mathcal{D}, \mathcal{E}', \mathcal{S}, \mathcal{O}_C', L^1,$ or $\mathcal{D}_{L^1}'$.

Step 1: Class $\mathcal{X}(\mathbb{S}^{m-1} \times \mathbb{R})$ of $\mathrm{R}f(\mathbf{u}, p)$

Hertle's results found [39, Th. 4.6, Cor. 4.8] that the Radon transform is the continuous injection

$$\mathrm{R} : \mathcal{X}(\mathbb{R}^m) \hookrightarrow \mathcal{X}(\mathbb{S}^{m-1} \times \mathbb{R}), \tag{A.2}$$

where $\mathcal{X} = \mathcal{D}, \mathcal{E}', \mathcal{S}, \mathcal{O}'_C, L^1$, or $\mathcal{D}'_{L^1}$; if $f \in \mathcal{X}(\mathbb{R}^m)$ then $\mathrm{R}f \in \mathcal{X}(\mathbb{S}^{m-1} \times \mathbb{R})$, which determines the second column. Our possible choice of the domain $\mathcal{X}$ is restricted to them.

Step 2: Class $\mathcal{B}(\mathbb{R})$ of $\mathcal{R}_\psi f(\mathbf{u}, \alpha, \beta)$ with respect to β

Fix $\alpha > 0$. Recall that $\mathcal{R}_\psi f(\mathbf{u}, \alpha, \beta) = \left(\mathrm{R}f(\mathbf{u}, \cdot) * \overline{\widetilde{\psi_\alpha}} \right)(\beta)$ in the sense of Schwartz distributions. By the nuclearity of $\mathcal{X}$ [33, §51], the kernel theorem

$$\mathcal{X}(\mathbb{S}^{m-1} \times \mathbb{R}) \cong \mathcal{X}(\mathbb{S}^{m-1}) \widehat{\otimes} \mathcal{X}(\mathbb{R}), \tag{A.3}$$

holds. Therefore, we can omit $\mathbf{u} \in \mathbb{S}^{m-1}$ in the considerations for $(\alpha, \beta) \in \mathbb{H}$. According to Schwartz's results shown in Table 3, for the convolution $g * \psi$ of $g \in \mathcal{X}(\mathbb{R})$ and $\psi \in \mathcal{Z}(\mathbb{R})$ to converge in $\mathcal{B}(\mathbb{R})$, we can assign the largest possible class $\mathcal{Z}$ for each $\mathcal{X}$ as in the third column. Note that for $\mathcal{X} = L^1$ we even assumed the continuity $\mathcal{Z} = L^p \cap C^0$, which is technically required in Step 3. Obviously for $\mathcal{Z} = \mathcal{D}', \mathcal{S}', L^p \cap C^0$, or $\mathcal{D}'_{L^p}$, if $\psi \in \mathcal{Z}(\mathbb{R})$ then $\psi_\alpha \in \mathcal{Z}(\mathbb{R})$. Therefore, we can determine the fourth column by evaluating $\mathcal{X} * \mathcal{Z}$ according to Table 3.

Step 3: Class $\mathcal{A}(\mathbb{H})$ of $\mathcal{R}_\psi f(\mathbf{u}, \alpha, \beta)$ with respect to (α, β)

Fix $\mathbf{u}_0 \in \mathbb{S}^{m-1}$ and assume $f \in \mathcal{X}(\mathbb{R}^m)$. Write $g(p) := \mathrm{R}f(\mathbf{u}_0, p)$ and

$$\mathcal{W}[\psi; g](\alpha, \beta) := \int_{\mathbb{R}} g(\alpha z + \beta) \overline{\psi(z)} \mathrm{d}z, \tag{A.4}$$

then $\mathcal{R}_\psi f(\mathbf{u}_0, \alpha, \beta) = \mathcal{W}[\psi; g](\alpha, \beta)$ for every $(\alpha, \beta) \in \mathbb{H}$. By the kernel theorem, $g \in \mathcal{X}(\mathbb{R})$.

Case 3a: ($\mathcal{X} = \mathcal{D}$ and $\mathcal{Z} = \mathcal{D}'$ then $\mathcal{B} = \mathcal{E}$ and $\mathcal{A} = \mathcal{E}$)

We begin by considering the case in the first row. Observe that

$$\partial_\alpha \mathcal{W}[\psi; g](\alpha, \beta) = \partial_\alpha \int_{\mathbb{R}} g(\alpha z + \beta) \overline{\psi(z)} \mathrm{d}z = \int_{\mathbb{R}} g'(\alpha z + \beta) \overline{z \cdot \psi(z)} \mathrm{d}z = \mathcal{W}[z \cdot \psi; g'](\alpha, \beta), \tag{A.5}$$

$$\partial_\beta \mathcal{W}[\psi; g](\alpha, \beta) = \partial_\beta \int_{\mathbb{R}} g(\alpha z + \beta) \overline{\psi(z)} \mathrm{d}z = \int_{\mathbb{R}} g'(\alpha z + \beta) \overline{\psi(z)} \mathrm{d}z = \mathcal{W}[\psi; g'](\alpha, \beta), \tag{A.6}$$

and thus that for every $k, \ell \in \mathbb{N}_0$,

$$\partial_\alpha^k \partial_\beta^\ell \mathcal{W}[\psi; g](\alpha, \beta) = \mathcal{W}[z^k \cdot \psi; g^{(k+\ell)}](\alpha, \beta). \tag{A.7}$$

Obviously if $g \in \mathcal{D}(\mathbb{R})$ and $\psi \in \mathcal{D}'(\mathbb{R})$ then $g^{(k+\ell)} \in \mathcal{D}(\mathbb{R})$ and $z^k \cdot \psi \in \mathcal{D}'(\mathbb{R})$, respectively, and thus $\partial_\alpha^k \partial_\beta^\ell \mathcal{W}[\psi; g](\alpha, \beta)$ exists at every $(\alpha, \beta) \in \mathbb{H}$. Therefore, we can conclude that if $g \in \mathcal{D}(\mathbb{R})$ and $\psi \in \mathcal{D}'(\mathbb{R})$ then $\mathcal{W}[\psi; g] \in \mathcal{E}(\mathbb{H})$.

Case 3b: ($\mathcal{X} = \mathcal{E}'$ and $\mathcal{Z} = \mathcal{D}'$ then $\mathcal{B} = \mathcal{D}'$ and $\mathcal{A} = \mathcal{D}'$)

Let $g \in \mathcal{E}'(\mathbb{R})$ and $\psi \in \mathcal{D}'(\mathbb{R})$. We show that $\mathcal{W}[\psi; g] \in \mathcal{D}'(\mathbb{H})$, that is, for every compact set $\mathrm{K} \subset \mathbb{H}$, there exists $N \in \mathbb{N}_0$ such that

$$\left| \int_{\mathrm{K}} \mathrm{T}(\alpha, \beta) \mathcal{W}[\psi; g](\alpha, \beta) \frac{\mathrm{d}\alpha \mathrm{d}\beta}{\alpha} \right| \lesssim \sum_{k, \ell \le N} \sup_{(\alpha, \beta) \in \mathbb{H}} |\partial_\alpha^k \partial_\beta^\ell \mathrm{T}(\alpha, \beta)|, \quad \forall \mathrm{T} \in \mathcal{D}(\mathrm{K}). \tag{A.8}$$

Fix an arbitrary compact set $K \subset \mathbb{H}$ and a smooth function $T \in \mathcal{D}(K)$, which is supported in K. Take two compact sets $A \subset \mathbb{R}_+$ and $B \subset \mathbb{R}$ such that $K \subset A \times B$. By the assumption that $g \in \mathcal{E}'(\mathbb{R})$ and $\psi \in \mathcal{D}'(\mathbb{R})$, there exist $k, \ell \in \mathbb{N}_0$ such that

$$\left| \int_{\mathbb{R}} u(z) g(z) \mathrm{d}z \right| \lesssim \sup_{z \in \mathrm{supp}\, g} |u^{(k)}(z)|, \quad \forall u \in \mathcal{E}(\mathbb{R}) \tag{A.9}$$

$$\left| \int_{\mathbb{R}} v(z) \overline{\psi(z)} \mathrm{d}z \right| \lesssim \sup_{z \in \mathbb{R}} |v^{(\ell)}(z)|, \quad \forall v \in \mathcal{D}(B). \tag{A.10}$$

Observe that for every fixed α, $T(\alpha, \cdot) * \widetilde{g} \in \mathcal{D}'(\mathbb{R})$. Then, by applying (A.9) and (A.10) incrementally,

$$\left| \int_{\mathbb{R}} T(\alpha, \beta) \int_{\mathbb{R}} g(\alpha z + \beta) \overline{\psi(z)} \mathrm{d}z \frac{\mathrm{d}\alpha \mathrm{d}\beta}{\alpha} \right| \leq \int_0^\infty \left| \int_{\mathbb{R}} \int_{\mathbb{R}} T(\alpha, \beta - \alpha z) \overline{\psi(z)} \mathrm{d}z \cdot g(\beta) \mathrm{d}\beta \right| \frac{\mathrm{d}\alpha}{\alpha} \tag{A.11}$$

$$\lesssim \int_0^\infty \sup_{\beta \in \mathrm{supp}\, g} \left| \int_{\mathbb{R}} \partial_\beta^k T(\alpha, \beta - \alpha z) \psi(z) \mathrm{d}z \right| \frac{\mathrm{d}\alpha}{\alpha} \tag{A.12}$$

$$\lesssim \int_0^\infty \sup_{\beta \in \mathrm{supp}\, g} \sup_z \left| \partial_\beta^{k+\ell} T(\alpha, \beta - \alpha z) \right| \alpha^{\ell-1} \mathrm{d}\alpha \tag{A.13}$$

$$= \int_A \sup_{\beta \in B} \left| \partial_\beta^{k+\ell} T(\alpha, \beta) \right| \alpha^{\ell-1} \mathrm{d}\alpha \tag{A.14}$$

$$\leq \sup_{(\alpha, \beta) \in K} \left| \partial_\beta^{k+\ell} T(\alpha, \beta) \right| \cdot \int_A \alpha^{\ell-1} \mathrm{d}\alpha, \tag{A.15}$$

where the third inequality follows by repeatedly applying $\partial_z [T(\alpha, \beta - \alpha z)] = (-\alpha) \partial_\beta T(\alpha, \beta - \alpha z)$; the fourth inequality follows by the compactness of the support of T. Thus, we conclude that $\mathcal{W}[\psi; g] \in \mathcal{D}'(\mathbb{H})$.

Case 3c: ($\mathcal{X} = \mathcal{S}$ and $\mathcal{Z} = \mathcal{S}'$ then $\mathcal{B} = \mathcal{O}_M$ and $\mathcal{A} = \mathcal{O}_M$)

Let $g \in \mathcal{S}(\mathbb{R})$ and $\psi \in \mathcal{S}'(\mathbb{R})$. Recall the case when $\mathcal{X} = \mathcal{D}$. Obviously, for every $k, \ell \in \mathbb{N}_0$, $g^{(k+\ell)} \in \mathcal{S}(\mathbb{R})$ and $z^k \cdot \psi \in \mathcal{S}'(\mathbb{R})$, respectively, which implies $\mathcal{W}[\psi; g] \in \mathcal{E}(\mathbb{H})$. Now we even show that $\mathcal{W}[\psi; g] \in \mathcal{O}_M(\mathbb{H})$, that is, for every $k, \ell \in \mathbb{N}_0$ there exist $s, t \in \mathbb{N}_0$ such that

$$\left| \partial_\alpha^k \partial_\beta^\ell \mathcal{W}[\psi; g](\alpha, \beta) \right| \lesssim (\alpha + 1/\alpha)^s (1 + \beta^2)^{t/2}. \tag{A.16}$$

Recall that by (A.7), we can regard $\partial_\alpha^k \partial_\beta^\ell \mathcal{W}[\psi; g](\alpha, \beta)$ as $\partial_\alpha^0 \partial_\beta^0 \mathcal{W}[\psi_0; g_0](\alpha, \beta)$, by setting $g_0 := g^{(k+\ell)} \in \mathcal{S}(\mathbb{R})$ and $\psi_0 := z^k \cdot \psi \in \mathcal{S}'(\mathbb{R})$. Henceforth we focus on the case when $k = \ell = 0$. Since $\psi \in \mathcal{S}'(\mathbb{R})$, there exists $N \in \mathbb{N}_0$ such that

$$\left| \int_{\mathbb{R}} u(z) \overline{\psi(z)} \mathrm{d}z \right| \lesssim \sum_{s, t \leq N} \sup_{z \in \mathbb{R}} |z^s u^{(t)}(z)|, \quad \forall u \in \mathcal{S}(\mathbb{R}). \tag{A.17}$$

By substituting $u(z) \leftarrow g(\alpha z + \beta)$, we have

$$\left| \int_{\mathbb{R}} g(\alpha z + \beta) \overline{\psi(z)} \mathrm{d}z \right| \lesssim \sum_{s,t \leq N} \sup_{z \in \mathbb{R}} |z^s \partial_z^t g(\alpha z + \beta)| \tag{A.18}$$

$$= \sum_{s,t \leq N} \sup_{p \in \mathbb{R}} \left| \left(\frac{p - \beta}{\alpha} \right)^s \alpha^t g^{(t)}(p) \right| \tag{A.19}$$

$$\lesssim \sum_{s,t \leq N} \alpha^{t-s} \beta^s \sup_{p \in \mathbb{R}} |p^s g^{(t)}(p)| \tag{A.20}$$

$$\lesssim (\alpha + 1/\alpha)^N (1 + \beta^2)^{N/2}, \tag{A.21}$$

where the second equation follows by substituting $p \leftarrow \alpha z + \beta$; the fourth inequality follows because every $\sup_p |p^s g^t(p)|$ is finite by assumption that $g \in \mathcal{S}(\mathbb{R})$. Therefore, we can conclude that if $g \in \mathcal{S}(\mathbb{R})$ and $\psi \in \mathcal{S}'(\mathbb{R})$ then $\mathcal{W}[\psi; g] \in \mathcal{O}_{\mathcal{M}}(\mathbb{H})$.

Case 3d: ($\mathcal{X} = \mathcal{O}_{\mathcal{C}}'$ and $\mathcal{Z} = \mathcal{S}'$ then $\mathcal{B} = \mathcal{S}'$ and $\mathcal{A} = \mathcal{S}'$)

Let $g \in \mathcal{O}_{\mathcal{C}}'(\mathbb{R})$ and $\psi \in \mathcal{S}'(\mathbb{R})$. We show that $\mathcal{W}[\psi; g] \in \mathcal{S}'(\mathbb{H})$, that is, there exists $N \in \mathbb{N}_0$ depending only on ψ and g such that

$$\left| \int_{\mathbb{H}} \mathrm{T}(\alpha, \beta) \mathcal{W}[\psi; g](\alpha, \beta) \frac{\mathrm{d}\alpha \mathrm{d}\beta}{\alpha} \right| \lesssim \sum_{s,t,k,\ell \leq N} \sup_{\alpha, \beta \in \mathbb{H}} \left| \mathrm{D}_{s,t}^{k,\ell} \mathrm{T}(\alpha, \beta) \right|, \quad \forall \mathrm{T} \in \mathcal{S}(\mathbb{H}) \tag{A.22}$$

where we defined

$$\mathrm{D}_{s,t}^{k,\ell} \mathrm{T}(\alpha, \beta) := (\alpha + 1/\alpha)^s (1 + \beta^2)^{t/2} \partial_\alpha^k \partial_\beta^\ell \mathrm{T}(\alpha, \beta). \tag{A.23}$$

Fix an arbitrary $\mathrm{T} \in \mathcal{S}(\mathbb{H})$. By the assumption that $\psi \in \mathcal{S}'(\mathbb{R})$, there exist $s, t \in \mathbb{N}_0$ such that

$$\left| \int_{\mathbb{R}} u(z) \overline{\psi(z)} \mathrm{d}z \right| \lesssim \sup_z |z^t u^{(s)}(z)|, \quad \forall u \in \mathcal{S}(\mathbb{R}). \tag{A.24}$$

Observe that for every fixed α, $\mathrm{T}(\alpha, \cdot) * \widetilde{g} \in \mathcal{S}(\mathbb{R})$. Then we can provide an estimate as below.

$$\left| \int_{\mathbb{H}} \mathrm{T}(\alpha, \beta) \int_{\mathbb{R}} g(\alpha z + \beta) \overline{\psi(z)} \mathrm{d}z \frac{\mathrm{d}\alpha \mathrm{d}\beta}{\alpha} \right| \leq \int_0^\infty \left| \int_{\mathbb{R}} \int_{\mathbb{R}} \mathrm{T}(\alpha, \beta) g(\alpha z + \beta) \mathrm{d}\beta \cdot \overline{\psi(z)} \mathrm{d}z \right| \frac{\mathrm{d}\alpha}{\alpha} \tag{A.25}$$

$$\lesssim \int_{\mathbb{R}} \sup_z \left| z^t \int_{\mathbb{R}} \mathrm{D}_{s,0}^{0,0} \mathrm{T}(\alpha, \beta) g^{(s)}(\alpha z + \beta) \mathrm{d}\beta \right| \frac{\mathrm{d}\alpha}{\alpha} \tag{A.26}$$

$$\lesssim \int_{\mathbb{R}} \sup_p \left| p^t \int_{\mathbb{R}} \mathrm{D}_{s+t,0}^{0,0} \mathrm{T}(\alpha, \beta) g^{(s)}(p + \beta) \mathrm{d}\beta \right| \frac{\mathrm{d}\alpha}{\alpha} \tag{A.27}$$

$$\leq \int_{\mathbb{R}} \int_{\mathbb{R}} \sup_p |p^t g^{(s)}(p + \beta)| \left| \mathrm{D}_{s+t,0}^{0,0} \mathrm{T}(\alpha, \beta) \right| \frac{\mathrm{d}\beta \mathrm{d}\alpha}{\alpha} \tag{A.28}$$

$$\lesssim \int_{\mathbb{R}} \int_{\mathbb{R}} \sup_p \left| (1 + |p + \beta|^2)^{t/2} g^{(s)}(p + \beta) \right| \left| \mathrm{D}_{s+t,t}^{0,0} \mathrm{T}(\alpha, \beta) \right| \frac{\mathrm{d}\beta \mathrm{d}\alpha}{\alpha} \tag{A.29}$$

$$\lesssim \int_{\mathbb{H}} \left| \mathrm{D}_{s+t,t}^{0,0} \mathrm{T}(\alpha, \beta) \right| \frac{\mathrm{d}\beta \mathrm{d}\alpha}{\alpha} \tag{A.30}$$

$$\leq \sup_{(\alpha,\beta)\in\mathbb{H}} \left|D^{0,0}_{s+t+\varepsilon,t+\delta}T(\alpha,\beta)\right| \int_{\mathbb{H}} (\alpha + 1/\alpha)^{-\varepsilon} (1+\beta^2)^{-\delta/2} \frac{\mathrm{d}\beta\mathrm{d}\alpha}{\alpha},$$

(A.31)

where the second inequality follows by repeatedly applying $\partial_z[g(\alpha z + \beta)] = \alpha \cdot g'(\alpha z + \beta)$ and $\alpha \lesssim \alpha + 1/\alpha$; the third inequality follows by changing the variable $p \leftarrow \alpha z$ and applying $(\alpha + 1/\alpha)^s \cdot \alpha^{-t} \lesssim (\alpha + 1/\alpha)^{s+t}$; the fifth inequality follows by applying $|p| \lesssim (1+p^2)^{1/2}$ and Peetre's inequality $1+p^2 \lesssim (1+\beta^2)(1+|p+\beta|^2)$; the sixth inequality follows by the assumption that $(1 + p^2)^{t/2}g(p)$ is bounded for any t; the last inequality follows by Hölder's inequality and the integral is convergent when $\varepsilon > 0$ and $\delta > 1$.

Case 3e: $(\mathcal{X} = L^1$ and $\mathcal{Z} = L^p \cap C^0$ then $\mathcal{B} = L^p \cap C^0$ and $\mathcal{A} = \mathcal{S}')$

Let $g \in L^1(\mathbb{R})$ and $\psi \in L^p \cap C^0(\mathbb{R})$. We show that $\mathcal{W}[g; \psi] \in \mathcal{S}'(\mathbb{H})$, that is, it has at most polynomial growth at infinity. Because ψ is continuous, $g * \psi$ is continuous. By Lusin's theorem, there exists a continuous function $g^\star$ such that $g^\star(x) = g(x)$ for almost every $x \in \mathbb{R}$; thus, by the continuity of $g * \psi$,

$$g^\star * \psi(x) = g * \psi(x), \quad \text{for every } x \in \mathbb{R}. \tag{A.32}$$

By the continuity and the integrability of $g^\star$ and ψ, there exist $s, t \in \mathbb{R}$ such that

$$|g^\star(x)| \lesssim (1 + x^2)^{-s/2}, \quad s > 1 \tag{A.33}$$

$$|\psi(x)| \lesssim (1 + x^2)^{-t/2}, \quad tp > 1 \tag{A.34}$$

Therefore,

$$\left| \int_{\mathbb{R}} g(x)\overline{\psi\left(\frac{x-\beta}{\alpha}\right)} \frac{1}{\alpha}\mathrm{d}x \right| \lesssim \left| \int_{\mathbb{R}} (1+x^2)^{-s/2}\left(1+\left(\frac{x-\beta}{\alpha}\right)^2\right)^{-t/2}\mathrm{d}x \right|\alpha^{-1} \tag{A.35}$$

$$\lesssim \left| \int_{\mathbb{R}} (1+x^2)^{-s/2}\left(1+(x-\beta)^2\right)^{-t/2}\mathrm{d}x \right|(1+\alpha^2)^{t/2}\alpha^{-1} \tag{A.36}$$

$$\lesssim (1+\beta^2)^{-\min(s,t)/2}(\alpha + 1/\alpha)^{t-1}, \tag{A.37}$$

which means $\mathcal{W}[\psi; g]$ is a locally integrable function that grows at most polynomially at infinity.

Note that if $(t - 1)p < m - 1$ then $\mathcal{W}[\psi; g] \in L^p(\mathbb{H}; \alpha^{-m}\mathrm{d}\alpha\mathrm{d}\beta)$, because $|\mathcal{W}[\psi; g](\alpha, \beta)|^p$ behaves as $\beta^{-\min(s,t)}\alpha^{(t-1)p}$ at infinity.

Case 3f: $(\mathcal{X} = \mathcal{D}'_{L^1}$ and $\mathcal{Z} = \mathcal{D}'_{L^p}$ then $\mathcal{B} = \mathcal{D}'_{L^p}$ and $\mathcal{A} = \mathcal{S}')$

Let $g \in \mathcal{D}'_{L^1}(\mathbb{R})$ and $\psi \in \mathcal{D}'_{L^p}$. We estimate (A.22). Fix an arbitrary $T \in \mathcal{S}(\mathbb{H})$. By the assumption that $\psi \in \mathcal{D}'_{L^p}(\mathbb{R})$, for every fixed α, $T(\alpha, \cdot) * \psi \in \mathcal{E} \cap L^p(\mathbb{R})$. Therefore, we can take $\psi^\star \in \mathcal{E} \cap L^p(\mathbb{R})$ such that $\psi^\star = \psi$ a.e. and $T(\alpha, \cdot) * \psi^\star = T(\alpha, \cdot) * \psi$. In the same way, we can take $g^\star \in \mathcal{E} \cap L^1(\mathbb{R})$ such that $g^\star = g$ almost everywhere and $T(\alpha, \cdot) * g^\star = T(\alpha, \cdot) * g$. Therefore, this case reduces to show that if $\mathcal{X} = L^1$ and $\mathcal{Z} = L^p$ then $\mathcal{A} = \mathcal{S}'$. This coincides with case 3e.

Step 4: Class $\mathcal{Y}(\mathbb{Y}^{m+1})$ of $\mathscr{R}_\psi f(\mathbf{u}, \alpha, \beta)$

The last column $(\mathcal{Y})$ is obtained by applying $\mathcal{Y}(\mathbb{Y}^{m+1}) = \mathcal{X}(\mathbb{S}^{m-1})\widehat{\otimes}\mathcal{A}(\mathbb{H})$. Recall that for $\mathbb{S}^{m-1}$, as it is compact, $\mathcal{D} = \mathcal{S} = \mathcal{O}_M = \mathcal{E}$ and $\mathcal{E}' = \mathcal{O}'_C = \mathcal{S}' = \mathcal{D}'_{L^p} = \mathcal{D}'$. Therefore, we have $\mathcal{Y}$ as in the last column of Table 4.

Appendix B. Proof of Theorem 5.4

Let $(\psi, \eta) \in \mathcal{S}(\mathbb{R}) \times \mathcal{S}'(\mathbb{R})$. Assume that $\widehat{\eta}$ is singular at 0. That is, there exists $k \in \mathbb{N}_0$ such that

$$\widehat{\eta}(\zeta) = \sum_{j=0}^{k} c_j \delta^{(j)}(\zeta), \quad \zeta \in \{0\}. \tag{B.1}$$

Assume there exists a neighborhood Ω of 0 such that $\widehat{\eta} \in C^0(\Omega \setminus \{0\})$. Note that the continuity implies local integrability. We show that ψ and η are admissible if and only if there exists $u \in \mathcal{O}_{\mathcal{M}}(\mathbb{R})$ such that

$$\Lambda^m u = \overline{\overline{\psi}} * \left(\eta - \sum_{j=0}^{k} c_j z^j \right), \quad \text{and} \quad \int_{\mathbb{R} \setminus \{0\}} \widehat{u}(\zeta) \mathrm{d}\zeta \neq 0. \tag{B.2}$$

Recall that the Fourier transform $\mathcal{O}_{\mathcal{M}}(\mathbb{R}) \to \mathcal{O}'_{\mathcal{C}}(\mathbb{R})$ is bijective. Thus, the action of $\widehat{u}$ on the indicator function $\mathbf{1}_{\mathbb{R} \setminus \{0\}}(\zeta)$ is always finite.

Sufficiency:

On $\Omega \setminus \{0\}$, $\widehat{\eta}$ coincides with a function. Thus the product $\overline{\widehat{\psi}(\zeta)} \widehat{\eta}(\zeta) |\zeta|^{-m}$ is defined in the sense of ordinary functions, and coincides with $\widehat{u}(\zeta)$. On $\mathbb{R} \setminus \Omega$, $|\zeta|^{-m}$ is in $\mathcal{O}_{\mathcal{M}}(\mathbb{R} \setminus \Omega)$. Thus, the product $\overline{\widehat{\psi}(\zeta)} \widehat{\eta}(\zeta) |\zeta|^{-m}$ is defined in the sense of distributions, which is associative because it contains at most one tempered distribution $(\mathcal{S} \cdot \mathcal{S}' \cdot \mathcal{O}_{\mathcal{M}})$, and reduces to $\widehat{u}(\zeta)$. Therefore,

$$\frac{K_{\psi,\eta}}{(2\pi)^{m-1}} = \left(\int_{\Omega \setminus \{0\}} + \int_{\mathbb{R} \setminus \Omega} \right) \widehat{u}(\zeta) \mathrm{d}\zeta, \tag{B.3}$$

which is finite by assumption.

Necessity:

Write $\Omega_0 := \Omega \cap [-1, 1]$ and $\Omega_1 := \mathbb{R} \setminus \Omega_0$. By the assumption that $\int_{\Omega_0 \setminus \{0\}} \overline{\widehat{\psi}(\zeta)} \widehat{\eta}(\zeta) |\zeta|^{-m} \mathrm{d}\zeta$ is absolutely convergent and $\widehat{\eta}$ is continuous in $\Omega_0 \setminus \{0\}$, there exists $\varepsilon > 0$ such that

$$\left| \overline{\widehat{\psi}(\zeta)} \widehat{\eta}(\zeta) \right| \lesssim |\zeta|^{m-1+\varepsilon}, \quad \zeta \in \Omega_0 \setminus \{0\}. \tag{B.4}$$

Therefore, there exists $v_0 \in L^1(\mathbb{R}) \cap C^0(\mathbb{R} \setminus \{0\})$ such that its restriction to $\Omega_0 \setminus \{0\}$ coincides with

$$\overline{\widehat{\psi}(\zeta)} \widehat{\eta}(\zeta) = |\zeta|^m v_0(\zeta), \quad \zeta \in \Omega_0 \setminus \{0\}. \tag{B.5}$$

By integrability and continuity, $v_0 \in L^\infty(\mathbb{R})$. In particular, both $\lim_{\zeta \to +0} v_0(\zeta)$ and $\lim_{\zeta \to -0} v_0(\zeta)$ are finite.

However, in Ω_1, $|\zeta|^{-m} \in \mathcal{O}_{\mathcal{M}}(\Omega_1)$. By the construction, $\overline{\widehat{\psi}} \cdot \widehat{\eta} \in \mathcal{O}'_{\mathcal{C}}(\mathbb{R})$. Thus, there exists $v_1 \in \mathcal{O}'_{\mathcal{C}}(\mathbb{R})$ such that

$$\overline{\widehat{\psi}(\zeta)} \widehat{\eta}(\zeta) = |\zeta|^m v_1(\zeta), \quad \zeta \in \Omega_1 \tag{B.6}$$

where the equality is in the sense of distribution.

Let

$$v := v_0 \cdot \mathbf{1}_{\Omega_0} + v_1 \cdot \mathbf{1}_{\Omega_1}. \tag{B.7}$$

Clearly, $v \in \mathcal{O}'_{\mathcal{C}}(\mathbb{R})$ because $v_0 \cdot \mathbf{1}_{\Omega_0} \in \mathcal{E}'(\mathbb{R})$ and $v_1 \cdot \mathbf{1}_{\Omega_1} \in \mathcal{O}'_{\mathcal{C}}(\mathbb{R})$. Therefore, there exists $u \in \mathcal{O}_{\mathcal{M}}(\mathbb{R})$ such that $\widehat{u} = v$ and

$$\overline{\widehat{\psi}(\zeta)} \widehat{\eta}(\zeta) = |\zeta|^m \widehat{u}(\zeta), \quad \zeta \in \mathbb{R} \setminus \{0\}. \tag{B.8}$$

By the admissibility condition,

$$\int_{\mathbb{R} \setminus \{0\}} \widehat{u}(\zeta) \mathrm{d}\zeta = \int_{\Omega_0 \setminus \{0\}} v_0(\zeta) \mathrm{d}\zeta + \int_{\Omega_1} v_1(\zeta) \mathrm{d}\zeta \neq 0. \tag{B.9}$$

In consideration of the singularity at 0, we have

$$\overline{\widehat{\psi}(\zeta)} \left(\widehat{\eta}(\zeta) - \sum_{j=0}^{k} c_j \delta^{(j)}(\zeta) \right) = |\zeta|^m \widehat{u}(\zeta), \quad \zeta \in \mathbb{R}. \tag{B.10}$$

By taking the Fourier inversion in the sense of distributions,

$$\left[\overline{\widetilde{\psi}} * \left(\eta - \sum_{j=0}^{k} c_j z^j \right) \right](z) = \Lambda^m u(z), \quad z \in \mathbb{R}. \tag{B.11}$$

Appendix C. Proof of Theorem 5.6

Let $f \in L^1(\mathbb{R}^m)$ satisfy $\widehat{f} \in L^1(\mathbb{R}^m)$ and $(\psi, \eta) \in \mathcal{S}(\mathbb{R}) \times \mathcal{S}'_0(\mathbb{R})$ be admissible. For simplicity, we rescale ψ to satisfy $K_{\psi,\eta} = 1$. Write

$$I(\mathbf{x}; \varepsilon, \delta) := \int_{\mathbb{S}^{m-1}} \int_{\varepsilon}^{\delta} \int_{\mathbb{R}} \mathscr{R}_{\psi} f(\mathbf{u}, \alpha, \mathbf{u} \cdot \mathbf{x} - \alpha z) \, \eta(z) \frac{\mathrm{d}z \mathrm{d}\alpha \mathrm{d}\mathbf{u}}{\alpha^m}. \tag{C.1}$$

We show that

$$\lim_{\substack{\delta \to \infty \\ \varepsilon \to 0}} I(\mathbf{x}; \varepsilon, \delta) = f(\mathbf{x}), \quad \text{a.e. } \mathbf{x} \in \mathbb{R}^m \tag{C.2}$$

and the equality holds at every continuous point of f.

By using the Fourier slice theorem in the sense of distribution,

$$\int_{\mathbb{R}} \mathscr{R}_{\psi} f(\mathbf{u}, \alpha, \beta - \alpha z) \, \eta(z) \mathrm{d}z = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{f}(\omega \mathbf{u}) \overline{\widehat{\psi}(\alpha\omega)} \widehat{\eta}(\alpha\omega) e^{i\omega\beta} \mathrm{d}\omega \tag{C.3}$$

$$= \frac{1}{2\pi} \int_{\mathbb{R} \setminus \{0\}} \widehat{f}(\omega \mathbf{u}) \widehat{u}(\alpha\omega) |\alpha\omega|^m e^{i\omega\beta} \mathrm{d}\omega, \tag{C.4}$$

where $|\zeta|^m \widehat{u}(\zeta) := \overline{\widehat{\psi}(\zeta)} \widehat{\eta}(\zeta)$ ($\zeta \neq 0$) is defined as in Theorem 5.4.

Then,

$$\int_\varepsilon^\delta \text{(C.4)} \frac{\mathrm{d}\alpha}{\alpha^m} = \frac{1}{2\pi} \int_{\mathbb{R}\setminus\{0\}} \int_\varepsilon^\delta \widehat{f}(\omega\mathbf{u})\widehat{u}(\alpha\omega)|\omega|^m e^{i\omega\beta}\mathrm{d}\alpha\mathrm{d}\omega \tag{C.5}$$

$$= \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\varepsilon \le \frac{\zeta}{\omega} \le \delta} \widehat{u}(\zeta)\widehat{f}(\omega\mathbf{u})e^{i\omega\beta}|\omega|^{m-1}\mathrm{d}\zeta\mathrm{d}\omega \tag{C.6}$$

$$= \frac{1}{2\pi} \int_0^\infty \int_{r\varepsilon \le |\zeta| \le r\delta} \widehat{u}(\zeta)\widehat{f}(\mathrm{sgn}(\zeta)r\mathbf{u}) \exp(\mathrm{sgn}(\zeta)ir\beta)r^{m-1}\mathrm{d}\zeta\mathrm{d}r, \tag{C.7}$$

where the second equation follows by changing the variable $\zeta \leftarrow \alpha\omega$ with $\alpha^{m-1}\mathrm{d}\alpha = |\omega|^{m-1}|\zeta|^{-m}\mathrm{d}\zeta$; the third equation follows by changing the variable $r \leftarrow |\omega|$ with $\mathrm{sgn}\,\omega = \mathrm{sgn}\,\zeta$. In the following, we substitute $\beta \leftarrow \mathbf{u} \cdot \mathbf{x}$. Observe that in $\int_{\mathbb{S}^{m-1}} \mathrm{d}\mathbf{u}$,

$$\int_{\mathbb{S}^{m-1}} \widehat{f}(-r\mathbf{u}) \exp(-ir\mathbf{u} \cdot \mathbf{x})\mathrm{d}\mathbf{u} = \int_{\mathbb{S}^{m-1}} \widehat{f}(r\mathbf{u}) \exp(ir\mathbf{u} \cdot \mathbf{x})\mathrm{d}\mathbf{u}. \tag{C.8}$$

Hence, we can omit $\mathrm{sgn}\,\zeta$.

Then, by substituting $\beta \leftarrow \mathbf{u} \cdot \mathbf{x}$ and changing the variable $\boldsymbol{\xi} \leftarrow r\mathbf{u}$,

$$I(\mathbf{x}; \varepsilon, \delta) = \int_{\mathbb{S}^{m-1}} \text{(C.7)} \,\mathrm{d}\mathbf{u} \tag{C.9}$$

$$= \frac{1}{2\pi} \int_{\mathbb{S}^{m-1}} \int_0^\infty \left[\int_{r\varepsilon \le |\zeta| \le r\delta} \widehat{u}(\zeta)\mathrm{d}\zeta \right] \widehat{f}(r\mathbf{u})e^{ir\mathbf{u}\cdot\mathbf{x}}r^{m-1}\mathrm{d}r\mathrm{d}\mathbf{u} \tag{C.10}$$

$$= \frac{1}{2\pi} \int_{\mathbb{R}^m} \left[\int_{\|\boldsymbol{\xi}\|\varepsilon \le |\zeta| \le \|\boldsymbol{\xi}\|\delta} \widehat{u}(\zeta)\mathrm{d}\zeta \right] \widehat{f}(\boldsymbol{\xi})e^{i\boldsymbol{\xi}\cdot\mathbf{x}}\mathrm{d}\zeta\mathrm{d}\boldsymbol{\xi}. \tag{C.11}$$

Recall that $\widehat{u} \in \mathcal{O}'_C(\mathbb{R})$; thus, its action is continuous. That is, the limits and the integral commute.

Therefore,

$$\mathscr{R}_\eta^\dagger \mathscr{R}_\psi f(\mathbf{x}) = \lim_{\substack{\delta\to\infty \\ \varepsilon\to 0}} I(\mathbf{x}; \varepsilon, \delta) \tag{C.12}$$

$$= \frac{1}{(2\pi)} \int_{\mathbb{R}^m} \left[\int_{\mathbb{R}\setminus\{0\}} \widehat{u}(\zeta)\mathrm{d}\zeta \right] \widehat{f}(\boldsymbol{\xi})e^{i\boldsymbol{\xi}\cdot\mathbf{x}}\mathrm{d}\boldsymbol{\xi} \tag{C.13}$$

$$= \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} \widehat{f}(\boldsymbol{\xi})e^{i\boldsymbol{\xi}\cdot\mathbf{x}}\mathrm{d}\boldsymbol{\xi} \tag{C.14}$$

$$= f(\mathbf{x}), \quad \text{a.e. } \mathbf{x} \in \mathbb{R}^m \tag{C.15}$$

where the last equation follows by the Fourier inversion formula, a consequence of which the equality holds at $\mathbf{x}_0$ if f is continuous at $\mathbf{x}_0$.

Appendix D. Proof of Theorem 5.7

Let $f \in L^1(\mathbb{R}^m)$ and $(\psi, \eta) \in \mathcal{S}(\mathbb{R}) \times \mathcal{S}'(\mathbb{R})$. Assume that there exists $u \in \mathcal{E} \cap L^1(\mathbb{R})$ that is real-valued, $\Lambda^m u = \overline{\psi} * \eta$ and $\int_{\mathbb{R}} \widehat{u}(\zeta) \mathrm{d}\zeta = -1$. Write

$$I(\mathbf{x}; \varepsilon, \delta) := \int_{\mathbb{S}^{m-1}} \int_{\varepsilon}^{\delta} \int_{\mathbb{R}} \mathscr{R}_\psi f (\mathbf{u}, \alpha, \mathbf{u} \cdot \mathbf{x} - \alpha z) \, \eta(z) \frac{\mathrm{d}z \mathrm{d}\alpha \mathrm{d}\mathbf{u}}{\alpha^m}. \tag{D.1}$$

We show that

$$\lim_{\substack{\delta \to \infty \\ \varepsilon \to 0}} I(\mathbf{x}; \varepsilon, \delta) = \mathrm{R}^* \Lambda^{m-1} \mathrm{R}(\mathbf{x}), \quad \text{a.e. } \mathbf{x} \in \mathbb{R}^m. \tag{D.2}$$

In the following we write $(\cdot)_\alpha(p) = (\cdot)(p/\alpha)/\alpha$. By using the convolution form,

$$\int_{\mathbb{R}} \mathscr{R}_\psi f (\mathbf{u}, \alpha, \beta - \alpha z) \, \eta(z) \mathrm{d}z = \left[\mathrm{R}f(\mathbf{u}, \cdot) * \left(\overline{\overline{\psi}} * \eta \right)_\alpha \right](\beta) \tag{D.3}$$

$$= \left[\mathrm{R}f(\mathbf{u}, \cdot) * (\Lambda^m u)_\alpha \right](\beta). \tag{D.4}$$

Observe that

$$\int_{\varepsilon}^{\delta} (\Lambda^m u)_\alpha(p) \frac{\mathrm{d}\alpha}{\alpha^m} = \Lambda^{m-1} \left[\int_{\varepsilon}^{\delta} (\Lambda u) \left(\frac{p}{\alpha} \right) \frac{\mathrm{d}\alpha}{\alpha^2} \right] \tag{D.5}$$

$$= \Lambda^{m-1} \left[\frac{1}{p} \int_{p/\delta}^{p/\varepsilon} (\Lambda u)(z) \mathrm{d}z \right] \tag{D.6}$$

$$= \Lambda^{m-1} \left[\frac{1}{p} \mathcal{H}u \left(\frac{p}{\varepsilon} \right) - \frac{1}{p} \mathcal{H}u \left(\frac{p}{\delta} \right) \right] \tag{D.7}$$

$$= \Lambda^{m-1} [k_\varepsilon(p) - k_\delta(p)], \tag{D.8}$$

where the first equality follows by repeatedly applying $(\Lambda u)_\alpha = \alpha \Lambda(u_\alpha)$; the second equality follows by substituting $z \leftarrow p/\alpha$; the fourth equality follows by defining

$$k(z) := \frac{1}{z} \mathcal{H}u(z) \quad \text{and} \quad k_\gamma(p) := \frac{1}{\gamma} k \left(\frac{p}{\gamma} \right) \quad \text{for } \gamma = \varepsilon, \delta. \tag{D.9}$$

Therefore, we have

$$\int_{\varepsilon}^{\delta} (\mathrm{D.4}) \frac{\mathrm{d}\alpha}{\alpha^m} = [\mathrm{R}f(\mathbf{u}, \cdot) * (\mathrm{D.8})](\beta) \tag{D.10}$$

$$= \left[\Lambda^{m-1} \mathrm{R}f(\mathbf{u}, \cdot) * (k_\varepsilon - k_\delta) \right](\beta). \tag{D.11}$$

We show that $k \in L^1 \cap L^\infty(\mathbb{R})$ and $\int_{\mathbb{R}} k(z) \mathrm{d}z = 1$. To begin with, $k \in L^1(\mathbb{R})$ because there exist $s, t > 0$ such that

$$|k(z)| \lesssim |z|^{-1+s}, \quad \text{as } |z| \to 0 \tag{D.12}$$

$$|k(z)| \lesssim |z|^{-1-t}, \quad \text{as } |z| \to \infty. \tag{D.13}$$

The first claim holds because u is real-valued and thus $\widehat{u}$ is odd, then

$$\mathcal{H}u(0) = \int_{\mathbb{R}} \operatorname{sgn}\zeta \cdot \widehat{u}(\zeta)\mathrm{d}\zeta \tag{D.14}$$

$$= \int_{(-\infty,0]} \widehat{u}(\zeta)\mathrm{d}\zeta - \int_{(0,\infty)} \widehat{u}(\zeta)\mathrm{d}\zeta \tag{D.15}$$

$$= 0. \tag{D.16}$$

The second claim holds because $u \in L^1(\mathbb{R})$ and thus u as well as $\mathcal{H}u$ decays at infinity. Then, by the continuity and the integrability of k, it is bounded. By the assumption that $\int_{\mathbb{R}} \widehat{u}(\zeta)\mathrm{d}\zeta = -1$,

$$\int_{\mathbb{R}} k(z)\mathrm{d}z = -\int_{\mathbb{R}} \frac{\mathcal{H}u(z)}{0-z}\mathrm{d}z \tag{D.17}$$

$$= -u(0) \tag{D.18}$$

$$= 1. \tag{D.19}$$

Write

$$J(\mathbf{u},p) := \Lambda^{m-1}\mathrm{R}f(\mathbf{u},p). \tag{D.20}$$

Because $k \in L^1(\mathbb{R})$ and $\int_{\mathbb{R}} k(z)\mathrm{d}z = 1$, k_ε is an approximation of the identity [43, III, Th. 2]. Then,

$$\lim_{\varepsilon\to 0} J(\mathbf{u},\cdot) * k_\varepsilon(p) = J(\mathbf{u},p), \quad \text{a.e. } (\mathbf{u},p) \in \mathbb{S}^{m-1} \times \mathbb{R}. \tag{D.21}$$

However, as $k \in L^\infty(\mathbb{R})$,

$$\|J * k_\delta\|_{L^\infty(\mathbb{S}^{m-1}\times\mathbb{R})} \leq \delta^{-1}\|J\|_{L^1(\mathbb{S}^{m-1}\times\mathbb{R})}\|k\|_{L^\infty(\mathbb{R})}, \tag{D.22}$$

and thus,

$$\lim_{\delta\to\infty} J(\mathbf{u},\cdot) * k_\delta(p) = 0, \quad \text{a.e. } (\mathbf{u},p) \in \mathbb{S}^{m-1} \times \mathbb{R}. \tag{D.23}$$

Because it is an approximation to the identity, $J * k_\gamma \in L^1(\mathbb{S}^{m-1} \times \mathbb{R})$ for $0 \leq \gamma$. Hence, there exists a maximal function $M(\mathbf{u},p)$ [43, III, Th. 2] such that

$$\sup_{0<\varepsilon} |(J(\mathbf{u},\cdot) * v_\varepsilon)(p)| \lesssim M(\mathbf{u},p). \tag{D.24}$$

Therefore, $|J(\mathbf{u},\cdot) * (v_\varepsilon - v_\delta)(\mathbf{u}\cdot\mathbf{x})|$ is uniformly integrable [35, Ex. 4.15.4] on $\mathbb{S}^{m-1}$. That is, if $\Omega \subset \mathbb{S}^{m-1}$ satisfies $\int_\Omega \mathrm{d}\mathbf{u} \leq A$ then

$$\int_\Omega |J(\mathbf{u},\cdot) * k_\gamma|(\mathbf{u}\cdot\mathbf{x})\mathrm{d}\mathbf{u} \lesssim A \sup_{\mathbf{u},p} |M(\mathbf{u},p)|, \quad \forall\gamma \geq 0. \tag{D.25}$$

Thus, by the Vitali convergence theorem, we have

$$\mathscr{R}_\eta^\dagger \mathscr{R}_\psi f(\mathbf{x}) = \lim_{\substack{\delta \to \infty \\ \varepsilon \to 0}} \int_{\mathbb{S}^{m-1}} [J(\mathbf{u}, \cdot) * (v_\varepsilon - v_\delta)](\mathbf{u} \cdot \mathbf{x}) \mathrm{d}\mathbf{u} \tag{D.26}$$

$$= \int_{\mathbb{S}^{m-1}} J(\mathbf{u}, \mathbf{u} \cdot \mathbf{x}) \mathrm{d}\mathbf{u}, \quad \text{a.e. } \mathbf{x} \in \mathbb{R}^m \tag{D.27}$$

$$= \mathrm{R}^* \Lambda^{m-1} \mathrm{R} f(\mathbf{x}). \tag{D.28}$$

Appendix E. Proof of Theorem 5.11

Let $f \in L^2(\mathbb{R}^m)$ and (ψ, η) be admissible with $K_{\psi,\eta} = 1$. Assume without loss of generality that (ψ, ψ) and (η, η) are self-admissible respectively. Write

$$I[f; (\varepsilon, \delta)](\mathbf{x}) := \int_{\mathbb{S}^{m-1}} \int_\varepsilon^\delta \int_\mathbb{R} \mathscr{R}_\psi f(\mathbf{u}, \alpha, \mathbf{u} \cdot \mathbf{x} - \alpha z) \, \eta(z) \frac{\mathrm{d}z \mathrm{d}\alpha \mathrm{d}\mathbf{u}}{\alpha^m}. \tag{E.1}$$

In the following we write $\Omega[\varepsilon, \delta] := \mathbb{S}^{m-1} \times [\mathbb{R}_+ \setminus (\varepsilon, \delta)] \times \mathbb{R} \subset \mathbb{Y}^{m+1}$. We show that

$$\lim_{\substack{\delta \to \infty \\ \varepsilon \to 0}} \left\| f - I[f; (\varepsilon, \delta)] \right\|_2 = 0. \tag{E.2}$$

Observe that

$$\left\| f - I[f; (\varepsilon, \delta)] \right\|_2 = \sup_{\|g\|_2 = 1} \left| (f - I[f; (\varepsilon, \delta)], g) \right| \tag{E.3}$$

$$= \sup_{\|g\|_2 = 1} \left| (\mathscr{R}_\psi f, \mathscr{R}_\eta g)_{\Omega[\varepsilon, \delta]} \right| \tag{E.4}$$

$$\le \sup_{\|g\|_2 = 1} \left| \mathscr{R}_\psi f \right|_{L^2(\Omega[\varepsilon, \delta])} \left\| \mathscr{R}_\eta g \right\|_{L^2(\mathbb{Y}^{m+1})} \tag{E.5}$$

$$= \sup_{\|g\|_2 = 1} \left| \mathscr{R}_\psi f \right|_{L^2(\Omega[\varepsilon, \delta])} \left\| g \right\|_2 \tag{E.6}$$

$$\to 0 \cdot 1, \quad \text{as } \varepsilon \to 0, \delta \to \infty \tag{E.7}$$

where the third inequality follows by the Schwartz inequality; the last limit follows by $\| \mathscr{R}_\psi f \|_{L^2(\Omega[\varepsilon, \delta])}$, which shrinks as the domain $\Omega[\varepsilon, \delta]$ tends to $\emptyset$.

Appendix F. Proofs of Example 6.4 and Example 6.10

Let $\sigma(z) := (1 + e^{-z})^{-1}$. Obviously $\sigma(z) \in \mathcal{E}(\mathbb{R})$.

Step 0: Derivatives of $\sigma(z)$.
For every $k \in \mathbb{N}$,

$$\sigma^{(k)}(z) = S_k(\sigma(z)), \tag{F.1}$$

where $S_k(z)$ is a polynomial defined by

$$S_k(z) := \begin{cases} z(1-z) & k = 1 \\ S'_{k-1}(z)S_1(z) & k > 1, \end{cases} \tag{F.2}$$

which is justified by induction on k.

Step 1: $\sigma, \tanh \in \mathcal{O}_\mathcal{M}(\mathbb{R})$.

Recall that $|\sigma(z)| \le 1$. Hence, for every $k \in \mathbb{N}$,

$$\left|\sigma^{(k)}(z)\right| = |S_k(\sigma(z))| \le \max_{z \in [0,1]} |S_k(z)| < \infty. \tag{F.3}$$

Therefore, every $k \in \mathbb{N}_0$, $\sigma^{(k)}(z)$ is bounded, which concludes $\sigma(z) \in \mathcal{O}_\mathcal{M}(\mathbb{R})$.

Hence, immediately $\tanh \in \mathcal{O}_\mathcal{M}(\mathbb{R})$ because

$$\tanh(z) = 2\sigma(2z) - 1. \tag{F.4}$$

Step 2: $\sigma^{(k)} \in \mathcal{S}(\mathbb{R})$, $k \in \mathbb{N}$.

Observe that

$$\sigma'(z) = (e^{z/2} + e^{-z/2})^{-2}. \tag{F.5}$$

Hence, $\sigma'(z)$ decays faster than any polynomial, which means $\sup_z |z^\ell \sigma'(z)| < \infty$ for any $\ell \in \mathbb{N}_0$. Then, for every $k, \ell \in \mathbb{N}_0$,

$$\sup_z \left|z^\ell \sigma^{(k+1)}(z)\right| = \sup_z \left|z^\ell S_{k+1}(\sigma(z))\right| \le \max_z |z^\ell \sigma'(z)| \cdot \max_z |S'_k(\sigma(z))| < \infty, \tag{F.6}$$

which concludes $\sigma' \in \mathcal{S}(\mathbb{R})$. Therefore, $\sigma^{(k)} \in \mathcal{S}(\mathbb{R})$ for every $k \in \mathbb{N}$.

Step 3: $\sigma^{(-1)} \in \mathcal{O}_\mathcal{M}(\mathbb{R})$.

Observe that

$$\sigma^{(-1)}(z) = \int_0^z \sigma(w)\mathrm{d}w. \tag{F.7}$$

Hence, it is already known that $[\sigma^{(-1)}]^{(k)} = \sigma^{(k-1)} \in \mathcal{O}_\mathcal{M}(\mathbb{R})$ for every $k \in \mathbb{N}$. We show that $\sigma^{(-1)}(z)$ has at most polynomial growth. Write

$$\rho(z) := \sigma^{(-1)}(z) - z_+. \tag{F.8}$$

Then $\rho(z)$ attains at 0 its maximum $\max_z \rho(z) = \log 2$, because $\rho'(z) < 0$ when $z > 0$ and $\rho'(z) > 0$ when $z < 0$. Therefore,

$$\left|\sigma^{(-1)}(z)\right| \le |\rho(z)| + |z_+| \le \log 2 + |z|, \tag{F.9}$$

which concludes $\sigma^{(-1)}(z) \in \mathcal{O}_\mathcal{M}$.

Step 4: $\eta = \sigma^{(k)}$ is admissible with $\psi = \Lambda^m \mathrm{G}$ when $k \in \mathbb{N}$ is positive and odd.

Recall that $\eta = \sigma^{(k)} \in \mathcal{S}(\mathbb{R})$. Hence, $\langle \widehat{\eta}, \widehat{\psi_0} \rangle = \langle \eta, \psi_0 \rangle$. Observe that if k is odd, then $\sigma^{(k)}$ is an odd function and thus $\langle \eta, \psi_0 \rangle = 0$. However, if k is even, then $\sigma^{(k)}$ is an even function and thus $\langle \eta, \psi_0 \rangle \ne 0$.

Step 5: σ and $\sigma^{(-1)}$ cannot be admissible with $\psi = \Lambda^m \mathrm{G}$.

This follows by Theorem 5.4, because both

$$\int_{\mathbb{R}} \left(\overline{\overline{\tilde{G}}} * \sigma \right)(z) \mathrm{d}z \quad \text{and} \quad \int_{\mathbb{R}} \left(\overline{\overline{\tilde{G}}} * \sigma^{(-1)} \right)(z) \mathrm{d}z, \tag{F.10}$$

diverge.

Step 6: σ and $\sigma^{(-1)}$ are admissible with $\psi = \Lambda^m G'$ and $\psi = \Lambda^m G''$, respectively.

Observe that both

$$u_0 := \overline{\overline{\tilde{G'}}} * \sigma = \overline{\overline{\tilde{G}}} * \sigma' \quad \text{and} \quad u_{-1} := \overline{\overline{\tilde{G''}}} * \sigma^{(-1)} = \overline{\overline{\tilde{G}}} * \sigma', \tag{F.11}$$

belong to $\mathcal{S}(\mathbb{R})$. Hence, u_0 and u_{-1} satisfy the sufficient condition in Theorem 5.4.

References

[1] N. Murata, An integral representation of functions using three-layered betworks and their approximation bounds, Neural Netw. 9 (6) (1996) 947–956, http://dx.doi.org/10.1016/0893-6080(96)00000-7.

[2] S. Sonoda, N. Murata, Sampling hidden parameters from oracle distribution, in: 24th Int. Conf. Artif. Neural Networks, in: LNCS, vol. 8681, Springer International Publishing, 2014, pp. 539–546.

[3] X. Glorot, A. Bordes, Y. Bengio, Deep sparse rectifier neural networks, in: 14th Int. Conf. Artif. Intell. Stat., vol. 15, JMLR W&CP, 2011, pp. 315–323.

[4] I. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, Y. Bengio, Maxout networks, in: 30th Int. Conf. Mach. Learn., vol. 28, JMLR W&CP, 2013, pp. 1319–1327.

[5] G.E. Dahl, T.N. Sainath, G.E. Hinton, Improving deep neural networks for LVCSR using rectified linear units and dropout, in: 2013 IEEE Int. Conf., Acoust. Speech Signal Process, IEEE, 2013, pp. 8609–8613.

[6] A.L. Maas, A.Y. Hannun, A.Y. Ng, Rectifier nonlinearities improve neural network acoustic models, in: ICML 2013 Work. Deep Learn., Audio, Speech, Lang. Process, 2013.

[7] K. Jarrett, K. Kavukcuoglu, M. Ranzato, Y. LeCun, What is the best multi-stage architecture for object recognition?, in: 2009 IEEE 12th Int. Conf. Comput. Vision, IEEE, 2009, pp. 2146–2153.

[8] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: Adv. Neural Inf. Process. Syst., vol. 25, Curran Associates, Inc., 2012, pp. 1097–1105.

[9] M.D. Zeiler, M. Ranzato, R. Monga, M.Z. Mao, K. Yang, Q. Le Viet, P. Nguyen, A.W. Senior, V. Vanhoucke, J. Dean, G.E. Hinton, On rectified linear units for speech processing, in: 2013 IEEE Int. Conf. Acoust. Speech Signal Process., IEEE, 2013, pp. 3517–3521.

[10] H. Mhaskar, C.A. Micchelli, Approximation by superposition of sigmoidal and radial basis functions, Adv. in Appl. Math. 13 (3) (1992) 350–373, http://dx.doi.org/10.1016/0196-8858(92)90016-P.

[11] M. Leshno, V.Y. Lin, A. Pinkus, S. Schocken, Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, Neural Netw. 6 (6) (1993) 861–867, http://dx.doi.org/10.1016/S0893-6080(05)80131-5.

[12] A. Pinkus, Approximation theory of the MLP model in neural networks, Acta Numer. 8 (1999) 143–195, http://dx.doi.org/10.1017/S0962492900002919.

[13] Y. Ito, Representation of functions by superpositions of a step or sigmoid function and their applications to neural network theory, Neural Netw. 4 (3) (1991) 385–394, http://dx.doi.org/10.1016/0893-6080(91)90075-G.

[14] P.C. Kainen, V. Kůrková, A. Vogt, A Sobolev-type upper bound for rates of approximation by linear combinations of Heaviside plane waves, J. Approx. Theory 147 (1) (2007) 1–10, http://dx.doi.org/10.1016/j.jat.2006.12.009.

[15] V. Kůrková, Complexity estimates based on integral transforms induced by computational units, Neural Netw. 33 (2012) 160–167, http://dx.doi.org/10.1016/j.neunet.2012.05.002.

[16] S.M. Carroll, B.W. Dickinson, Construction of neural nets using the Radon transform, in: Int. Jt. Conf. Neural Networks, vol. 1, IEEE, 1989, pp. 607–611.

[17] S. Helgason, Integral Geometry and Radon Transforms, Springer-Verlag, New York, 2011.

[18] B. Irie, S. Miyake, Capabilities of three-layered perceptrons, in: IEEE Int. Conf. Neural Networks, IEEE, 1988, pp. 641–648.

[19] K.-I. Funahashi, On the approximate realization of continuous mappings by neural networks, Neural Netw. 2 (3) (1989) 183–192, http://dx.doi.org/10.1016/0893-6080(89)90003-8.

[20] L.K. Jones, A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training, Ann. Statist. 20 (1) (1992) 608–613, http://dx.doi.org/10.1214/aos/1176348546.

[21] A.R. Barron, Universal approximation bounds for superpositions of a sigmoidal function, IEEE Trans. Inform. Theory 39 (3) (1993) 930–945, http://dx.doi.org/10.1109/18.256500.

[22] P.C. Kainen, V. Kůrková, M. Sanguineti, Approximating multivariable functions by feedforward neural nets, in: Handb. Neural Inf. Process., vol. 49, Springer, Berlin, Heidelberg, 2013, pp. 143–181.

[23] E.J. Candès, Harmonic analysis of neural networks, Appl. Comput. Harmon. Anal. 6 (2) (1999) 197–218, http://dx.doi.org/10.1006/acha.1998.0248.

[24] E.J. Candès, Ridgelets: theory and applications, Ph.D. thesis, Standford University, 1998.

[25] B. Rubin, The Calderón reproducing formula, windowed X-ray transforms, and radon transforms in L^p-spaces, J. Fourier Anal. Appl. 4 (2) (1998) 175–197, http://dx.doi.org/10.1007/BF02475988.

[26] D.L. Donoho, Tight frames of k-plane ridgelets and the problem of representing objects that are smooth away from d-dimensional singularities in $\mathbb{R}^n$, Proc. Natl. Acad. Sci. USA 96 (5) (1999) 1828–1833, http://dx.doi.org/10.1073/pnas.96.5.1828.

[27] D.L. Donoho, Ridge functions and orthonormal ridgelets, J. Approx. Theory 111 (2) (2001) 143–179, http://dx.doi.org/10.1006/jath.2001.3568.

[28] J.-L. Starck, F. Murtagh, J.M. Fadili, The ridgelet and curvelet transforms, in: Sparse Image Signal Process. Wavelets, Curvelets, Morphol. Divers., Cambridge University Press, 2010, pp. 89–118.

[29] B. Rubin, Convolution backprojection method for the k-plane transform, and Calderón's identity for ridgelet transforms, Appl. Comput. Harmon. Anal. 16 (3) (2004) 231–242, http://dx.doi.org/10.1016/j.acha.2004.03.003.

[30] S. Kostadinova, S. Pilipović, K. Saneva, J. Vindas, The ridgelet transform of distributions, Integral Transforms Spec. Funct. 25 (5) (2014) 344–358, http://dx.doi.org/10.1080/10652469.2013.853057.

[31] S. Kostadinova, S. Pilipović, K. Saneva, J. Vindas, The ridgelet transform and quasiasymptotic behavior of distributions, Oper. Theory Adv. Appl. 245 (2015) 185–197, http://dx.doi.org/10.1007/978-3-319-14618-8_13.

[32] L. Schwartz, Théorie des Distributions, nouvelle edition, Hermann, Paris, 1966.

[33] F. Trèves, Tological Vector Spaces, Distributions and Kernels, Academic Press, 1967.

[34] W. Rudin, Functional Analysis, 2nd edition, Higher Mathematics Series, McGraw–Hill Education, 1991.

[35] H. Brezis, Functional Analysis, Sobolev Spaces and Partial Differential Equations, 1st edition, Universitext, Springer-Verlag, New York, 2011.

[36] K. Yosida, Functional Analysis, 6th edition, Springer-Verlag, Berlin, Heidelberg, 1995.

[37] W. Yuan, W. Sickel, D. Yang, Morrey and Campanato Meet Besov, Lizorkin and Triebel, Lecture Notes in Mathematics, Springer, Berlin, Heidelberg, 2010.

[38] M. Holschneider, Wavelets: An Analysis Tool, Oxford Mathematical Monographs, The Clarendon Press, 1995.

[39] A. Hertle, Continuity of the radon transform and its inverse on Euclidean space, Math. Z. 184 (2) (1983) 165–192, http://dx.doi.org/10.1007/BF01252856.

[40] L. Grafakos, Classical Fourier Analysis, 2nd edition, Graduate Texts in Mathematics, Springer, New York, 2008.

[41] I.M. Gel'fand, G.E. Shilov, Properties and Operations, Generalized Functions, vol. 1, Academic Press, New York, 1964.

[42] L.A. Shepp, B.F. Logan, The Fourier reconstruction of a head section, IEEE Trans. Nucl. Sci. 21 (3) (1974) 21–43, http://dx.doi.org/10.1109/TNS.1974.6499235.

[43] E.M. Stein, Singular Integrals and Differentiability Properties of Functions, Princeton Mathematical Series (PMS), Princeton University Press, 1970.