A Continuous Transform for Localized Ridgelets

Joseph Shenouda
Department of Electrical and Computer Engineering,
University of Wisconsin–Madison
Madison, WI 53706
Email: jshenouda@wisc.edu

Rahul Parhi
Biomedical Imaging Group
École polytechnique fédérale de Lausanne
CH-1015 Lausanne, Switzerland
Email: rahul.parhi@epfl.ch

Robert D. Nowak
Department of Electrical and Computer Engineering,
University of Wisconsin–Madison
Madison, WI 53706
Email: rdnowak@wisc.edu

***Abstract*—We develop a new continuous wavelet-like transform for localized ridgelets. Contrary to the classical ridgelets (which are not local), this new dictionary exhibits desirable decay so that all the atoms lie in $L^2(\mathbb{R}^d)$. Furthermore, each localized ridgelet atom is itself a superposition of continuously many classical ridgelets. Our construction hinges on a careful wavelet analysis in the Radon domain, different than the usual Radon-domain wavelet analysis found in the study of classical ridgelets. This is crucial in ensuring the locality of our new, localized ridgelet atoms. We prove a continuous transform and inversion formula for this new dictionary. Finally, due to the locality of these atoms, we conjecture that this new dictionary is better conditioned than the system of non-local ridge functions used ubiquitously in modern neural networks.**

***Index Terms*—Localization, Radon transform, ridgelets, wavelets.**

I. INTRODUCTION

A ridge function is a function which maps $\mathbb{R}^d \to \mathbb{R}$ that can be written in the form

$$\boldsymbol{x} \mapsto r(\boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{x}), \tag{1}$$

where $r : \mathbb{R} \to \mathbb{R}$ is referred to as the *ridge profile* and $\boldsymbol{\alpha} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$ is referred to as the *ridge direction*. A ridge function is, in essence, a univariate function which is extended outward in d dimensions. This is due to the fact that such a function is constant along the hyperplanes $\boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{x} = c$, where $c \in \mathbb{R}$ [21]. Ridge functions are pervasive in mathematics, science, and engineering. For example,

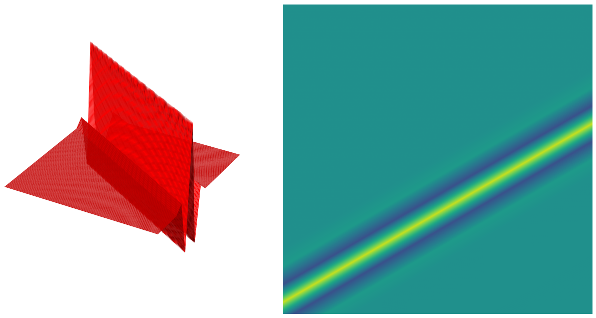
- plane waves are ridge functions with time-varying profiles and often arise as solutions to partial differential equations [9];
- the Fourier inversion formula provides a method to represent functions as a superposition of the ridge functions $\boldsymbol{x} \mapsto e^{\mathrm{j}\boldsymbol{\omega}^{\mathsf{T}}\boldsymbol{x}}$, $\boldsymbol{\omega} \in \mathbb{R}^d$ (complex exponentials);
- a neural network neuron with weights $\boldsymbol{w} \in \mathbb{R}^d$ and bias $b \in \mathbb{R}$ takes the form $\boldsymbol{x} \mapsto \sigma(\boldsymbol{w}^{\mathsf{T}}\boldsymbol{x} - b)$, where $\sigma : \mathbb{R} \to \mathbb{R}$ is the activation function.

Ridge functions are intimately tied to the Radon transform. This observation goes back to classical work regarding the representation of solutions to PDEs as superpositions of plane waves, in which the PDEs are analyzed in the Radon domain [9], [6]. The term "ridge function" was coined in [11] in the context of computerized tomography (CT), in which images are reconstructed from their Radon transform via ridge functions.

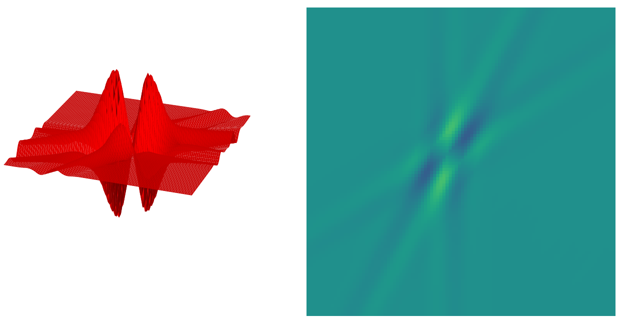
Recently, the study of shallow neural networks has led to researchers being interested in constructive approaches to representing multivariate functions with continuously many ridge functions (neurons). Early approaches are based on the continuous ridgelet transform, which corresponds to a univariate continuous wavelet transform in the Radon domain [2], [3], [13], [10], [22], [23]. Modern approaches are based on sparsity in the Radon domain [15], [17], [18], [19], [20].

Although ubiquitous, especially in modern neural networks, continuously indexed dictionaries of ridge functions are poorly conditioned due to the non-locality of ridge functions. In fact, ridge functions do not decay and therefore cannot lie in $L^p(\mathbb{R}^d)$ for any $1 \leq p < \infty$. Classical ridgelets are ridge functions with univariate wavelet profiles. In this paper, we revisit the continuous ridgelet transform from first principles. Our main contribution is the construction of a new dictionary of *localized ridgelets* (see (21)) with an accompanying continuous transform and inversion formula (Theorem 2) which proves that any function in $L^2(\mathbb{R}^d)$ can be represented as a continuous superposition of our localized ridgelets.

Contrary to classical ridgelet atoms our localized ridgelet atoms exhibit desirable decay and all lie in $L^2(\mathbb{R}^d)$. Remarkably, our localized ridgelets are themselves a continuous superposition of classical ridgelets (cf. (21)). The key idea behind our construction is a careful wavelet analysis in the (half-filtered) Radon domain inspired by Donoho's construction of a bivariate orthonormal basis via a wavelet analysis in the Radon domain [5]. Moreover, due to the Radon transform in the construction of the localized ridgelets, this dictionary can efficiently represent functions with singularities along hyperplanes, which is not possible with standard multivariate wavelets. We also note that a discretely indexed version of our localized ridgelets was recently explored in [16, Chapter 5]. Here, the authors use a similar wavelet analysis in the Radon domain to construct a tight frame of localized ridgelets for $L^2(\mathbb{R}^d)$. A comparison of a classical ridgelet with one of our localized ridgelets appears in Fig. 1. We also visualize how each parameter in our localized ridgelets effects the orientation of these atoms in Fig. 2.



(a) Classical ridgelet



(b) Our new, localized ridgelet

Fig. 1: Classical ridgelets based on a 2nd order B-spline wavelet vs. our new, localized ridgelet based on 2nd order B-spline wavelet and the Haar wavelet on the sphere. The left plots are surface plots of the functions and the right plots are heatmaps.

II. Mathematical Preliminaries

The Fourier transform of $f \in L^1(\mathbb{R})$ is given by

$$\widehat{f}(\omega) = \mathscr{F}\{f\}(\omega) = \int_{\mathbb{R}} f(t) e^{-\mathrm{j}\omega t} \, \mathrm{d}t, \quad \omega \in \mathbb{R}, \tag{2}$$

where $\mathrm{j}^2 = -1$. The Fourier transform is extended to functions in $L^2(\mathbb{R})$ in the usual way. Let $\psi \in L^2(\mathbb{R})$ be a real-valued admissible wavelet, i.e.,

$$C_\psi \coloneqq \int_0^\infty \frac{|\widehat{\psi}(\omega)|^2}{|\omega|} \, \mathrm{d}\omega = \int_{-\infty}^0 \frac{|\widehat{\psi}(\omega)|^2}{|\omega|} \, \mathrm{d}\omega < \infty. \tag{3}$$

We generate a continuously indexed wavelet dictionary by dilating and translating ψ:

$$\psi_{a,b}(t) = a^{-1/2} \psi\left(\frac{t-b}{a}\right), \quad a > 0, b \in \mathbb{R}. \tag{4}$$

The variable a plays the role of the *scale* of the wavelet and the variable b plays the role of the *location* of the wavelet. The continuous wavelet transform (CWT) on $\mathbb{R}$ is an operator that maps $L^2(\mathbb{R}) \to L^2(\mathbb{R}_{>0} \times \mathbb{R})$ given by $f \mapsto ((a, b) \mapsto \langle f, \psi_{a,b} \rangle)$. This transform is invertible via the reconstruction formula

$$f = \int_0^\infty \int_{-\infty}^\infty \langle f, \psi_{a,b} \rangle \psi_{a,b}(\cdot) \, \mathrm{d}b \, v(a) \, \mathrm{d}a, \tag{5}$$

where $v(a) = C_\psi^{-1} a^{-2}$ and where the equality is understood in the weak sense, i.e., taking the inner product of both sides in the above display with any $g \in L^2(\mathbb{R})$. We refer the reader to [4, Chapter 2] for more details about the continuous wavelet transform. Here and in the remainder of the paper, $\cdot$ will be used as a placeholder for dummy variables taking values in $\mathbb{R}$ (or $\mathbb{R}^d$).

There have been many works developing a spherical version of the continuous wavelet transform dating back to the 1990s [1], [7]. The primary difficulties that arise when working on the sphere in $\mathbb{R}^d$, denoted by

$$\mathbb{S}^{d-1} = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 = 1\}, \tag{6}$$

is that there is no dilation operator on $\mathbb{S}^{d-1}$ and that the rotation group on $\mathbb{S}^{d-1}$, denoted by $\mathrm{SO}(d)$, is much more complicated than the translation group on Euclidean space [14, p. 531]. Nevertheless, there exist continuously indexed dictionaries of real-valued spherical wavelets $\varphi_{s,\mathbf{R}} \in L^2(\mathbb{S}^{d-1})$. Morally, $s > 0$ plays the role of the scale of the wavelet and $\mathbf{R} \in \mathrm{SO}(d)$ plays the role of the location of the wavelet. Under some admissibility assumptions [8], these spherical wavelets admit a CWT that maps $L^2(\mathbb{S}^{d-1}) \to L^2(\mathbb{R}_{>0} \times \mathrm{SO}(d))$ given by $g \mapsto ((s, \mathbf{R}) \mapsto \langle g, \varphi_{s,\mathbf{R}} \rangle)$ which is invertible via the reconstruction formula

$$g = \int_0^\infty \int_{\mathrm{SO}(d)} \langle f, \varphi_{s,\mathbf{R}} \rangle \varphi_{s,\mathbf{R}}(\star) \, \mathrm{d}\sigma(\mathbf{R}) w(s) \, \mathrm{d}s, \tag{7}$$

which holds in the weak sense, where σ is the Haar measure on $\mathrm{SO}(d)$. Here, w is a positive weight function which plays the same role as v in (5). Here and in the remainder of the paper, $\star$ will be used as a placeholder for dummy variables taking values in $\mathbb{S}^{d-1}$.

The Radon transform is well-defined for any $f \in L^1(\mathbb{R}^d)$ and is given by the integral

$$\mathscr{R}\{f\}(\boldsymbol{\alpha}, t) = \int_{\boldsymbol{\alpha}^\perp} f(\boldsymbol{x} + \boldsymbol{\alpha} t) \, \mathrm{d}\boldsymbol{x}, \quad (\boldsymbol{\alpha}, t) \in \mathbb{S}^{d-1} \times \mathbb{R}, \tag{8}$$

where $\boldsymbol{\alpha}^\perp \coloneqq \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{\alpha}^\mathsf{T} \boldsymbol{x} = 0\}$. The dual transform (the "backprojection" operator) of $F \in L^\infty(\mathbb{S}^{d-1} \times \mathbb{R})$ is given by

$$\mathscr{R}^*\{F\}(\boldsymbol{x}) = \int_{\mathbb{S}^{d-1}} F(\boldsymbol{\alpha}, \boldsymbol{\alpha}^\mathsf{T} \boldsymbol{x}) \, \mathrm{d}\boldsymbol{\alpha}, \quad \boldsymbol{x} \in \mathbb{R}^d, \tag{9}$$

where $\mathrm{d}\boldsymbol{\alpha}$ denotes integration against the Haar measure on $\mathbb{S}^{d-1}$. For sufficiently nice functions $f : \mathbb{R}^d \to \mathbb{R}$ (e.g., Schwartz), the Radon transform is inverible by the filtered backprojection operator. That is to say, $\mathscr{R}^* \mathrm{K} \mathscr{R} f = f$, where K is a univariate filtering operator (in the t variable) defined in the Fourier domain by $\widehat{\mathrm{K} h}(\omega) = c_d |\omega|^{d-1} \widehat{h}(\omega)$. Here, $c_d = 1/(2(2\pi)^{d-1})$ is a constant that often arises when working with the Radon transform. A key property of the Radon transform and related operators is a bijective L^2-isometry from $L^2(\mathbb{R}^d)$ to $L^2_{\mathrm{even}}(\mathbb{S}^{d-1} \times \mathbb{R})$, the subspace of even L^2-functions on $\mathbb{S}^{d-1} \times \mathbb{R}$ [12]. The even subspace arises since the Radon transform is always even. This bijective isometry is given by

$$\mathrm{K}^{\frac{1}{2}} \mathscr{R} : L^2(\mathbb{R}^d) \to L^2_{\mathrm{even}}(\mathbb{S}^{d-1} \times \mathbb{R}) \tag{10}$$

with inverse

$$\mathscr{R}^* \, \mathrm{K}^{\frac{1}{2}} : L^2_{\mathrm{even}}(\mathbb{S}^{d-1} \times \mathbb{R}) \to L^2(\mathbb{R}^d). \tag{11}$$

Here, $\mathrm{K}^{\frac{1}{2}}$ is defined in the Fourier domain by $\widehat{\mathrm{K}^{\frac{1}{2}} g}(\omega) = \sqrt{c_d}|\omega|^{\frac{d-1}{2}}\widehat{g}(\omega)$. In other words, there is an L^2-isometry between the spatial domain and the half-filtered Radon domain.

III. A New Continuous Transform

In this section we construct a new continuously indexed dictionary where the atoms take the form of localized ridgelets. Unlike the classical ridgelet atoms, these new, localized, ridgelet atoms are in $L^2(\mathbb{R}^d)$. Furthermore, this localized ridgelet dictionary admits a continuous transform and reconstruction formula akin to (5) and (7).

The key idea will be to construct tensor product wavelet dictionary on $\mathbb{S}^{d-1} \times \mathbb{R}$ and pull the tensor product dictionary through the L^2-isometry $\mathscr{R}^* \, \mathrm{K}^{\frac{1}{2}}$ which maps $L^2_{\mathrm{even}}(\mathbb{S}^{d-1} \times \mathbb{R})$ to $L^2(\mathbb{R}^d)$ inspired by [5].

To begin, define the tensor product function

$$W_{a,b,s,\mathbf{R}} \coloneqq \mathrm{P}_{\mathrm{even}}\{\varphi_{s,\mathbf{R}} \otimes \psi_{a,b}\}, \tag{12}$$

where $\mathrm{P}_{\mathrm{even}}$ is the even projector. Written explicitly, we have that

$$W_{a,b,s,\mathbf{R}}(\boldsymbol{\alpha}, t) = \frac{\varphi_{s,\mathbf{R}}(\boldsymbol{\alpha})\psi_{a,b}(t) + \varphi_{s,\mathbf{R}}(-\boldsymbol{\alpha})\psi_{a,b}(-t)}{2} \tag{13}$$

To simplify notation, let $\lambda = (a, b, s, \mathbf{R})$ denote the parameters of of the tensor-product wavelet (12) and let $\Lambda = \mathbb{R}_{>0} \times \mathbb{R} \times \mathbb{R}_{>0} \times \mathrm{SO}(d)$ denote the parameter space.

Lemma 1. *Every $F \in L^2_{\mathrm{even}}(\mathbb{S}^{d-1} \times \mathbb{R})$ admits decomposition*

$$F(\star, \cdot) = \int_\Lambda \langle F, W_\lambda \rangle W_\lambda(\star, \cdot) \, \mathrm{d}\mu(\lambda), \tag{14}$$

where $\mathrm{d}\mu(\lambda) = \mathrm{d}b \, v(a) \, \mathrm{d}a \, \mathrm{d}\sigma(\mathbf{R}) \, w(s) \, \mathrm{d}s$ and the equality holds in the weak sense.

Proof. First, define the system $V_\lambda \coloneqq \varphi_{s,\mathbf{R}} \otimes \psi^{a,b}$, $\lambda \in \Lambda$. Next, define $\mathrm{d}\zeta(a, b) \coloneqq \mathrm{d}b \, v(a) \, \mathrm{d}a$ and $\mathrm{d}\gamma(\mathbf{R}, s) \coloneqq \mathrm{d}\sigma(\mathbf{R}) \, w(s) \, \mathrm{d}s$.

Given $H \in L^2(\mathbb{S}^{d-1} \times \mathbb{R})$, we that

$$\begin{aligned} &\int_\Lambda \langle H, V_\lambda \rangle V_\lambda(\star, \cdot) \, \mathrm{d}\mu(\lambda) \\ &= \int_\Lambda \langle H, \varphi_{s,\mathbf{R}} \otimes \psi_{a,b} \rangle \varphi_{s,\mathbf{R}}(\star)\psi_{a,b}(\cdot) \, \mathrm{d}\mu(\lambda) \\ &= \int_\Lambda \langle \langle H(\star, \cdot), \psi_{a,b} \rangle, \varphi_{s,\mathbf{R}} \rangle \varphi_{s,\mathbf{R}}(\star)\psi_{a,b}(\cdot) \, \mathrm{d}\mu(\lambda) \\ &= \int \left\langle \int \langle H(\star, \cdot), \psi_{a,b} \rangle \psi_{a,b} \, \mathrm{d}\zeta(a, b), \varphi_{s,\mathbf{R}} \right\rangle \varphi_{s,\mathbf{R}}(\star) \, \mathrm{d}\gamma(s, \mathbf{R}), \end{aligned} \tag{15}$$

where we note that interchanging integrals is allowed by Fubini's theorem since all functions are L^2-functions. If we take the inner product of the above display with any $G \in L^2(\mathbb{S}^{d-1} \times \mathbb{R})$ and interchange integrals accordingly, we find that

$$H(\star, \cdot) = \int_\Lambda \langle H, V_\lambda \rangle V_\lambda(\star, \cdot) \, \mathrm{d}\mu(\lambda) \tag{16}$$

in the weak sense. Since $L^2_{\mathrm{even}}(\mathbb{S}^{d-1} \times \mathbb{R}) \subset L^2(\mathbb{S}^{d-1} \times \mathbb{R})$, for any $F, G \in L^2_{\mathrm{even}}(\mathbb{S}^{d-1} \times \mathbb{R})$ we have

$$\begin{aligned} \langle F, G \rangle &= \left\langle \int_\Lambda \langle F, V_\lambda \rangle V_\lambda(\star, \cdot) \, \mathrm{d}\mu(\lambda), G \right\rangle \\ &= \left\langle \int_\Lambda \langle \mathrm{P}_{\mathrm{even}} \, F, V_\lambda \rangle V_\lambda(\star, \cdot) \, \mathrm{d}\mu(\lambda), G \right\rangle \\ &= \left\langle \int_\Lambda \langle F, \mathrm{P}_{\mathrm{even}} \, V_\lambda \rangle V_\lambda(\star, \cdot) \, \mathrm{d}\mu(\lambda), G \right\rangle \\ &= \left\langle \int_\Lambda \langle F, W_\lambda \rangle V_\lambda(\star, \cdot) \, \mathrm{d}\mu(\lambda), G \right\rangle \\ &= \left\langle \int_\Lambda \langle F, W_\lambda \rangle V_\lambda(\star, \cdot) \, \mathrm{d}\mu(\lambda), \mathrm{P}_{\mathrm{even}} \, G \right\rangle \\ &= \left\langle \mathrm{P}_{\mathrm{even}} \left\{ \int_\Lambda \langle F, W_\lambda \rangle V_\lambda(\star, \cdot) \, \mathrm{d}\mu(\lambda) \right\}, G \right\rangle \\ &= \left\langle \int_\Lambda \langle F, W_\lambda \rangle \, \mathrm{P}_{\mathrm{even}}\{V_\lambda\}(\star, \cdot) \, \mathrm{d}\mu(\lambda), G \right\rangle \\ &= \left\langle \int_\Lambda \langle F, W_\lambda \rangle W_\lambda(\star, \cdot) \, \mathrm{d}\mu(\lambda), G \right\rangle, \end{aligned} \tag{17}$$

where we took advantage of the fact that $\mathrm{P}_{\mathrm{even}}$ is an orthoprojector and therefore self-adjoint. □

In the remainder of this section, suppose that ψ satisfies the stronger admissibility condition

$$\int_0^\infty \frac{|\widehat{\psi}(\omega)|^2}{|\omega|^d} \, \mathrm{d}\omega = \int_{-\infty}^0 \frac{|\widehat{\psi}(\omega)|^2}{|\omega|^d} \, \mathrm{d}\omega < \infty. \tag{18}$$

This stronger condition ensures that ψ has $\lceil \frac{d-1}{2} \rceil + 1$ vanishing moments. Using the L^2-isometry $\mathscr{R}^* \, \mathrm{K}^{\frac{1}{2}}$, we can define the following continuously indexed dictionary on $\mathbb{R}^d$.

$$\rho_\lambda \coloneqq \mathscr{R}^* \, \mathrm{K}^{\frac{1}{2}} \, W_\lambda, \quad \lambda \in \Lambda. \tag{19}$$

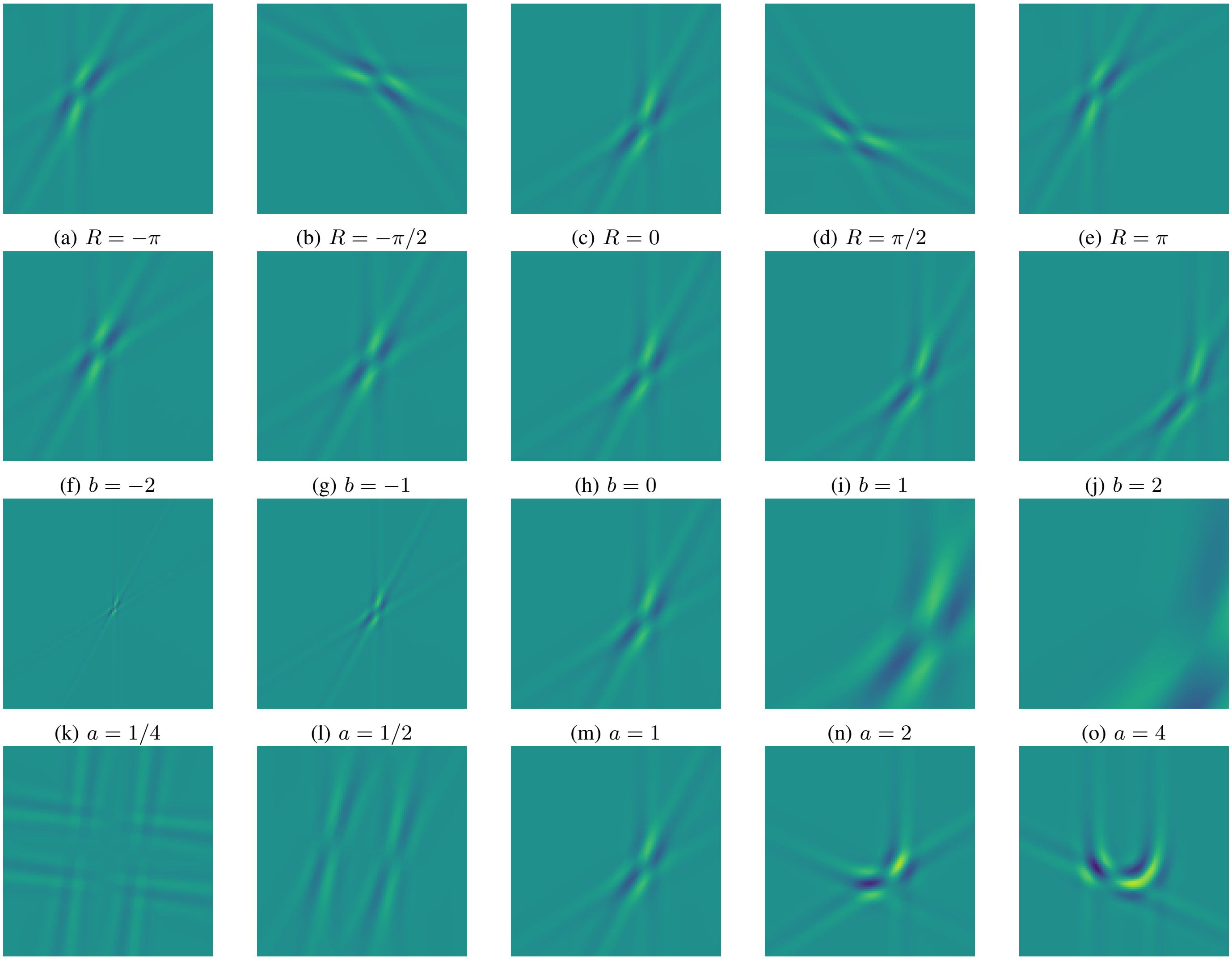
Since $W_\lambda \in L^2_{\mathrm{even}}(\mathbb{S}^{d-1} \times \mathbb{R})$, we have that $\rho_\lambda \in L^2(\mathbb{R}^d)$. The stronger admissibility condition on ψ ensures $\mathrm{K}^{\frac{1}{2}} \, W_\lambda$ is a bounded function so that the application of $\mathscr{R}^*$ is well-defined. If we define $\Psi_{a,b} \coloneqq \mathrm{K}^{\frac{1}{2}} \, \psi_{a,b}$, we have

$$(\mathrm{K}^{\frac{1}{2}} \, W_\lambda)(\boldsymbol{\alpha}, t) = \frac{\varphi_{s,\mathbf{R}}(\boldsymbol{\alpha})\Psi_{a,b}(t) + \varphi_{s,\mathbf{R}}(\boldsymbol{\alpha})\Psi_{a,b}(-t)}{2}, \tag{20}$$

where we used the fact that $\mathrm{K}^{\frac{1}{2}}$ commutes with reflections. Since ψ satisfies the stronger admissibility condition, $\Psi_{a,b}$ is itself a univariate wavelet dictionary. Thus,

$$\begin{aligned} \rho_\lambda(\boldsymbol{x}) &= (\mathscr{R}^* \, \mathrm{K}^{\frac{1}{2}} \, W_\lambda)(\boldsymbol{x}) \\ &= \frac{1}{2} \int_{\mathbb{S}^{d-1}} \varphi_{s,\mathbf{R}}(\boldsymbol{\alpha})\Psi_{a,b}(\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{x}) \\ &\qquad + \varphi_{s,\mathbf{R}}(-\boldsymbol{\alpha})\Psi_{a,b}(-\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{x}) \, \mathrm{d}\boldsymbol{\alpha} \\ &= \int_{\mathbb{S}^{d-1}} \Psi_{a,b}(\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{x})\varphi_{s,\mathbf{R}}(\boldsymbol{\alpha}) \, \mathrm{d}\boldsymbol{\alpha}. \end{aligned} \tag{21}$$

We refer to these atoms as *localized* ridgelets since each ρ_λ is constructed via a "classical" ridgelet $\boldsymbol{x} \mapsto \Psi_{a,b}(\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{x})$ combined with a localization procedure via integrating the classical ridgelet against a spherical wavelet with a particular



(a) $R = -\pi$ (b) $R = -\pi/2$ (c) $R = 0$ (d) $R = \pi/2$ (e) $R = \pi$

(f) $b = -2$ (g) $b = -1$ (h) $b = 0$ (i) $b = 1$ (j) $b = 2$

(k) $a = 1/4$ (l) $a = 1/2$ (m) $a = 1$ (n) $a = 2$ (o) $a = 4$

(p) $s = 1/4$ (q) $s = 1/2$ (r) $s = 1$ (s) $s = 2$ (t) $s = 4$

Fig. 2: Here we demonstrate how each parameter effects the orients and scales the atoms in our newly proposed dictionary. The top row corresponds to angular rotations. In the second row we vary the translation parameter for the wavelet on $\mathbb{R}$. In the third row we show the effects of different scaling on the wavelet on $\mathbb{R}$. Finally, in the fourth row we vary the scaling for the wavelet on the sphere. For each row we vary only one parameter while fixing the others to their canonical values ($a = 1, b = 0, s = 1, R = 0$). The localized ridgelet uses the 2nd order B-spline wavelet on $\mathbb{R}^2$ and a periodic Haar wavelet on the sphere.

location and scale. This localization procedure ensures that the localized ridgelet atoms decay quickly and all lie in $L^2(\mathbb{R}^d)$.

For any $f \in L^2(\mathbb{R}^d)$ we can now define its continuous localized ridgelet transform (CLRT) as the function $\lambda \mapsto \langle f, \rho_\lambda \rangle$. The CLRT is an operator that maps $L^2(\mathbb{R}^d) \to L^2(\Lambda)$. Furthermore, this operator is invertible by the reconstruction formula given in Theorem 2.

Theorem 2. *Every $f \in L^2(\mathbb{R}^d)$ admits the decomposition,*

$$f = \int_\Lambda \langle f, \rho_\lambda \rangle \rho_\lambda(\cdot)\, \mathrm{d}\mu(\lambda) \tag{22}$$

where the equality holds in the weak sense.

Proof. Given $f \in L^2(\mathbb{R}^d)$, $\mathrm{K}^{\frac{1}{2}} \mathscr{R} f \in L^2_{\text{even}}(\mathbb{S}^{d-1} \times \mathbb{R})$. By Lemma 1 we have the decomposition

$$\begin{aligned} \mathrm{K}^{\frac{1}{2}} \mathscr{R} f &= \int_\Lambda \langle \mathrm{K}^{\frac{1}{2}} \mathscr{R} f, W_\lambda \rangle W_\lambda(\star, \cdot)\, \mathrm{d}\mu(\lambda) \\ &= \int_\Lambda \langle f, \mathscr{R}^* \mathrm{K}^{\frac{1}{2}} W_\lambda \rangle W_\lambda(\star, \cdot)\, \mathrm{d}\mu(\lambda) \\ &= \int_\Lambda \langle f, \rho_\lambda \rangle W_\lambda(\star, \cdot)\, \mathrm{d}\mu(\lambda) \end{aligned} \tag{23}$$

in the weak sense. The application of $\mathscr{R}^* \mathrm{K}^{\frac{1}{2}}$ yields

$$f = \int_\Lambda \langle f, \rho_\lambda \rangle \rho_\lambda(\cdot)\, \mathrm{d}\mu(\lambda) \tag{24}$$

in the weak sense. □

IV. Visualization of the Localized Ridgelets

In this section we visally investigate our new system of atoms. In Fig. 2 we isolate each parameter of the localized ridgelet atoms and show how they influence their orientation and shape. We see that the parameters associated with the wavelet on $\mathbb{R}$ behave predictably. These parameters essentially apply a scaling or translation to the localized ridgelet. In contrast, the parameters of the spherical wavelet apply nontrivial transformations to the atoms.

V. Conclusion and Future Work

In this work, we developed a new continuous wavelet-like transform. The atoms of this new dictionary take the form of localized ridgelets. This system is the continuously indexed version of the recently developed discretely indexed tight frame of localized ridgelets for $L^2(\mathbb{R}^d)$ [16, Chapter 5]. Future work will be devoted to investigating the use of these localized ridgelets as a replacement for the standard neurons in neural networks. We conjecture that the superior conditioning of the localized dictionary will lead to fast convergence of gradient-based methods for data-fitting problems.

Acknowledgment

This work was supported by the ONR MURI grant N00014-20-1-2787, the NSF grants DMS-2134140 and DMS-2023239, the AFOSR/AFRL grant FA9550-18-1-0166.

References

[1] J.-P. Antoine and P. Vandergheynst, "Wavelets on the 2-sphere: A group-theoretical approach," *Applied and Computational Harmonic Analysis*, vol. 7, no. 3, pp. 262–291, 1999.

[2] E. J. Candès, "Ridgelets: theory and applications," Ph.D. dissertation, Stanford University Stanford, 1998.

[3] E. J. Candès, "Harmonic analysis of neural networks," *Applied and Computational Harmonic Analysis*, vol. 6, no. 2, pp. 197–218, 1999.

[4] I. Daubechies, *Ten lectures on wavelets*. SIAM, 1992.

[5] D. L. Donoho, "Orthonormal ridgelets and linear singularities," *SIAM Journal on Mathematical Analysis*, vol. 31, no. 5, pp. 1062–1099, 2000.

[6] L. C. Evans, *Partial Differential Equations*, ser. Graduate studies in mathematics. American Mathematical Society, 2010.

[7] M. Holschneider, "Continuous wavelet transforms on the sphere," *Journal of Mathematical Physics*, vol. 37, no. 8, pp. 4156–4165, 1996.

[8] I. Iglewska-Nowak, "Continuous wavelet transforms on n-dimensional spheres," *Applied and Computational Harmonic Analysis*, vol. 39, no. 2, pp. 248–276, 2015.

[9] F. John, *Plane Waves and Spherical Means Applied to Partial Differential Equations*, ser. Interscience tracts in pure and applied mathematics. Interscience publishers, 1981.

[10] S. Kostadinova, S. Pilipović, K. Saneva, and J. Vindas, "The ridgelet transform of distributions," *Integral Transforms and Special Functions*, vol. 25, no. 5, pp. 344–358, 2014.

[11] B. F. Logan and L. A. Shepp, "Optimal reconstruction of a function from its projections," *Duke Mathematical Journal*, vol. 42, no. 4, pp. 645–659, 1975.

[12] D. Ludwig, "The Radon transform on Euclidean space," *Communications on Pure and Applied Mathematics*, vol. 19, no. 1, pp. 49–81, 1966.

[13] N. Murata, "An integral representation of functions using three-layered networks and their approximation bounds," *Neural Networks*, vol. 9, no. 6, pp. 947–956, 1996.

[14] F. Narcowich, P. Petrushev, and J. Ward, "Decomposition of Besov and Triebel–Lizorkin spaces on the sphere," *Journal of Functional Analysis*, vol. 238, no. 2, pp. 530–564, 2006.

[15] G. Ongie, R. Willett, D. Soudry, and N. Srebro, "A function space view of bounded norm infinite width ReLU nets: The multivariate case," in *International Conference on Learning Representations*, 2020.

[16] R. Parhi, "On ridge splines, neural networks, and variational problems in Radon-domain BV spaces," Ph.D. dissertation, The University of Wisconsin-Madison, 2022.

[17] R. Parhi and R. D. Nowak, "Banach space representer theorems for neural networks and ridge splines," *Journal of Machine Learning Research*, vol. 22, no. 43, pp. 1–40, 2021.

[18] R. Parhi and R. D. Nowak, "What kinds of functions do deep neural networks learn? Insights from variational spline theory," *SIAM Journal on Mathematics of Data Science*, vol. 4, no. 2, pp. 464–489, 2022.

[19] R. Parhi and R. D. Nowak, "Deep learning meets sparse regularization: A signal processing perspective," *arXiv preprint arXiv:2301.09554*, 2023.

[20] R. Parhi and R. D. Nowak, "Near-minimax optimal estimation with shallow ReLU neural networks," *IEEE Transactions on Information Theory*, vol. 69, no. 2, pp. 1125–1140, 2023.

[21] A. Pinkus, *Ridge Functions*, ser. Cambridge Tracts in Mathematics. Cambridge University Press, 2015.

[22] B. Rubin, "The Calderón reproducing formula, windowed X-ray transforms, and Radon transforms in L^p-spaces," *Journal of Fourier Analysis and Applications*, vol. 4, no. 2, pp. 175–197, 1998.

[23] S. Sonoda and N. Murata, "Neural network with unbounded activation functions is universal approximator," *Applied and Computational Harmonic Analysis*, vol. 43, no. 2, pp. 233–268, 2017.