Journal of Machine Learning Research 24 (2023) 1-33 Submitted 3/22; Revised 11/22; Published 1/23

Ridges, Neural Networks, and the Radon Transform

Michael Unser MICHAEL.UNSER@EPFL.CH
Biomedical Imaging Group
École polytechnique fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland

Editor: Lorenzo Rosasco

Abstract

A ridge is a function that is characterized by a one-dimensional profile (activation) and a multidimensional direction vector. Ridges appear in the theory of neural networks as functional descriptors of the effect of a neuron, with the direction vector being encoded in the linear weights. In this paper, we investigate properties of the Radon transform in relation to ridges and to the characterization of neural networks. We introduce a broad category of hyper-spherical Banach subspaces (including the relevant subspace of measures) over which the back-projection operator is invertible. We also give conditions under which the back-projection operator is extendable to the full parent space with its null space being identifiable as a Banach complement. Starting from first principles, we then characterize the sampling functionals that are in the range of the filtered Radon transform. Next, we extend the definition of ridges for any distributional profile and determine their (filtered) Radon transform in full generality. Finally, we apply our formalism to clarify and simplify some of the results and proofs on the optimality of ReLU networks that have appeared in the literature.

1. Introduction

A ridge is a multidimensional function $\boldsymbol{x} \mapsto r(\boldsymbol{w}^\mathsf{T}\boldsymbol{x})$ from $\mathbb{R}^d \to \mathbb{R}$ that is characterized by a 1D profile $r : \mathbb{R} \to \mathbb{R}$ and a weight vector $\boldsymbol{w} \in \mathbb{R}^d \backslash \{\boldsymbol{0}\}$ (Pinkus, 2015). Ridges are ubiquitous in mathematics and engineering. Most significantly, the elementary unit (neuron) in a neural network is a function of the form $f_k(\boldsymbol{x}) = \sigma(\boldsymbol{w}_k^\mathsf{T}\boldsymbol{x} - t_k)$, which is a ridge with a shifted profile $r = \sigma(\cdot - t_k)$, where $\sigma : \mathbb{R} \to \mathbb{R}$ is the activation function and where $t_k \in \mathbb{R}$ (bias) and $\boldsymbol{w}_k \in \mathbb{R}^d$ (linear weights) are the trainable parameters of the kth neuron (Bishop, 2006). Variants of the universal-approximation theorem ensure that any continuous function can be approximated as closely as desired by a weighted sum of ridges with a fixed activation under mild conditions on σ (Cybenko, 1989; Hornik et al., 1989; Barron, 1993).

Ridges are also intimately tied to the Radon transform (Logan and Shepp, 1975; Madych, 1990) under the condition that the weight vector $\boldsymbol{w}$ has a unit norm, so that $\boldsymbol{w} \in \mathbb{S}^{d-1}$ where $\mathbb{S}^{d-1}$ is the unit sphere in $\mathbb{R}^d$ whose generic elements will be denoted by $\boldsymbol{\xi}$. This connection is exploited in the ridgelet transform, which provides a wavelet-like representation of functions where the basis elements are ridges (Murata, 1996; Rubin, 1998; Candès, 1999; Candès and Donoho, 1999; Kostadinova et al., 2014). The expansion of a function in terms of ridgelets is a precusor to sparse signal approximation. There, the idea is to represent a function by a linear combination of a small number of atoms taken within a dictionary (Elad, 2010; Foucart and Rauhut, 2013). This paradigm, which is the basis for compressed sensing (Donoho, 2006; Candès and Romberg, 2007), has been adapted to shallow neural networks by considering a dictionary that consists of a continuum of neurons. Mathematically, this can be implemented through the integral representation (infinite-width neural network)

$$f(\boldsymbol{x}) = \int_{\mathbb{R}\times\mathbb{S}^{d-1}} \sigma(\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x} - t)\mathrm{d}\mu(t,\boldsymbol{\xi}), \tag{1}$$

where μ is a measure on $\mathbb{R}\times\mathbb{S}^{d-1}$ (hyper-spherical domain). This model is fitted to data subject to a penalty on the total-variation norm of μ (Bach, 2017). Remarkably, this infinite-dimensional convex optimization problem results in sparse minimizers of the form $\mu = \sum_{k=1}^{K} a_k \delta_{\boldsymbol{z}_k}$ with $\boldsymbol{z}_k = (t_k, \boldsymbol{\xi}_k) \in \mathbb{R}\times\mathbb{S}^{d-1}$, which then map into standard two-layer neural networks (Bach, 2017). Interestingly, we can relate (1) to the Radon transform by identifying μ as the (generalized) function g_μ (with $\mathrm{d}\mu(t,\boldsymbol{\xi}) = g_\mu(t,\boldsymbol{\xi})\mathrm{d}t\mathrm{d}\boldsymbol{\xi}$) and by rewriting the integral as

$$f(\boldsymbol{x}) = \int_{\mathbb{S}^{d-1}} \left(\int_{\mathbb{R}} \sigma(\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x} - t) g_\mu(t,\boldsymbol{\xi})\mathrm{d}t \right) \mathrm{d}\boldsymbol{\xi} = \mathrm{R}^*\mathrm{L}_{\mathrm{rad}}\{g_\mu\}(\boldsymbol{x}), \tag{2}$$

where R^* (the adjoint of the Radon transform) is the back-projection operator of computer tomography (Natterer, 1984). Our "radial" operator $\mathrm{L}_{\mathrm{rad}} : g_\mu \mapsto \sigma \circledast g_\mu$ on the right-hand side of (2) implements the Radon-domain convolution with σ along the variable t.

While the synthesis approach to the learning problem proposed by Bach (2017) is insightful, there is a strong incentive to make the connection with regularization theory in direct analogy with the classical theory of learning that relies on reproducing-kernel Hilbert spaces (Wahba, 1990; Poggio and Girosi, 1990; Schölkopf et al., 1997, 2001; Alvarez et al., 2012; Unser, 2021). This is feasible provided that the linear relation between f and g_μ expressed by (2) be one-to-one. This requires that the operators $\mathrm{L}_{\mathrm{rad}}$ and R^* in (2) be both invertible. Ongie et al. (2020) made an important step in that direction by showing that ReLU networks are minimizers of a Radon-domain total-variation norm that involves the Laplacian of f. Their optimality result was then generalized by Parhi and Nowak (2021) who considered a broader class of differential operators inspired by spline theory (Unser et al., 2017). The leading idea there is that the operator $\mathrm{L}_{\mathrm{rad}}$ in (2) should implement some variant of an nth-order integrator, with $\sigma = \mathrm{ReLU}$ being the solution for $n = 2$. Such an $\mathrm{L}_{\mathrm{rad}}$ can be formally inverted by applying an nth-order partial derivative (e.g., $\mathrm{L}_{\mathrm{rad}}^{-1} = \partial_t^n$), which motivates the use of the latter as (filtered) Radon-domain regularization operator.

The proposed spline-based approach to the inversion of (2) is elegant and intuitively appealing. However, the formulation and resolution of the corresponding optimization problem requires special care because the underlying function spaces have a nontrivial kernel (null space) that needs to be factored out. The latter statement applies not only to the regularization operator (e.g., Laplacian and/or Radon-domain radial derivatives) but also to R^*, which is an aspect that has been overlooked. While there is a rich theory on the invertibility of the Radon transform (Helgason, 2011; Rubin, 1998; Boman and Lindskog, 2009; Ramm and Katsevich, 2020), there are comparatively fewer—and not as strong—results on the invertibility of R^*, the problem being that this operator has a huge null space (Ludwig, 1966). The primary spaces on which R^* is known to be injective, and hence invertible, are

- $\mathcal{S}(\mathbb{P}^d)$ (the even part of Schwartz' hyper-spherical—or Radon-domain—test functions) (Solmon, 1987);
- $L_{\infty,c}(\mathbb{P}^d)$ (the bounded even functions of compact support) (Ramm, 1996);
- $\mathcal{S}'_{\mathrm{Liz}}(\mathbb{P}^d)$ (the even Lizorkin distributions) (Kostadinova et al., 2014).

The space $\mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$ of Lizorkin distributions, which is the topological dual of $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ (the subspace of Schwartz functions that are orthogonal to all polynomials), is especially attractive in that context. Indeed, the Radon transform being an homeomorphism from $\mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$ onto $\mathcal{S}'_{\mathrm{Liz}}(\mathbb{P}^d)$, the inversion process is straightforward (Kostadinova et al., 2014). Lizorkin distributions also interact very nicely with the Laplace operator, which makes then well suited to the investigation of fractional integrals (Samko et al., 1993) and of wavelets (Saneva and Vindas, 2010). The Lizorkin framework, however, has one basic limitation. The underlying objects—Lirzorkin distributions—are abstract entities, with $\mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$ being isomorphic to the quotient space $\mathcal{S}'(\mathbb{R}^d)/\mathcal{P}$, where $\mathcal{S}'(\mathbb{R}^d)$ and $\mathcal{P}$ are the spaces of tempered distributions and polynomials, respectively. Thus, Lizorkin distributions are generally identifiable only modulo some polynomial. Fortunately, this is not a problem when dealing with ordinary functions $f \in L_p(\mathbb{R}^d)$ since $L_p(\mathbb{R}^d)$ is continuously embedded in $\mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$ for any $p > 1$ (Samko, 1982). This implies that the Lizorkin distribution $f + \mathcal{P} \in \mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$ has a unique "concrete" representer $f \in L_p(\mathbb{R}^d)$, which amounts to simply setting the polynomial to zero. The situation, however, is not as clearcut for functions and ridge profiles that exhibit polynomial growth at infinity. To offer insights on the nature of the problem, let us consider three distinct neuronal units $f_1(x) = (x - t_k)_+$ (ReLU activation), $f_2(x) = \frac{1}{2}|x - t_k|$, and $f_3(x) = x + (x - t_k)_+$ (ReLU with skip connection), which are all valid representers of the same Lizorkin distribution $f = f_i + \mathcal{P} \in \mathcal{S}'_{\mathrm{Liz}}(\mathbb{R})$ for $i = 1, 2, 3$ (since the f_i's only differ by a first-order polynomial). Suppose that a theoretical argument can be made concerning the optimality of the announced $f \in \mathcal{S}'_{\mathrm{Liz}}(\mathbb{R})$. The practical difficulty then is to map this result into a concrete architecture. Should the choice be one of the f_i, if any? The least we can say is that the convenient rule of "setting the polynomial to zero" is not applicable here because it is unclear what the underlying polynomial truly is. This intrinsic ambiguity jeopardizes some of the conclusions regarding the connection between ReLU neural networks, ridge splines, and the Radon transform that have been reported in the literature (Sonoda and Murata, 2017; Parhi and Nowak, 2021). We are in the opinion that adjustments are needed.

In this paper, we revisit the topic and extend the existing formulation so that it can handle arbitrary ridge profiles, without (polynomial) ambiguity. Our four primary contributions are as follows.

- A detailed investigation of the invertibility of the back-projection operator for a broad family of Radon-compatible Banach subspaces $\mathcal{X}'_{\mathrm{Rad}} \subset \mathcal{X}'$, where $\mathcal{X}'$ is the topological dual of some generic hyper-spherical parent space $\mathcal{X}$.
- A constructive characterization of the extreme points of the space of Radon-compatible hyper-spherical measures.
- The extension of ridges to distributional profiles $r \in \mathcal{S}'(\mathbb{R})$ and the determination of their Radon transform.

- The application of the formalism to the investigation of a functional-optimization problem that results in solutions that are parameterized by ReLU neural networks. Our contribution there is to clarify the analysis of Parhi and Nowak (2021) and to provide a characterization of the full solution set.

The paper is organized as follows: We start with notations and mathematical preliminaries in Section 2. In particular, we recall the main properties of the classical Radon transform and its adjoint, and show how to extend them to tempered distributions by duality. In Section 3, we develop a formulation that leads to the identification of a generic family of Radon-compatible Banach spaces over which the back-projection operator is invertible. Our results are summarized in Theorem 8, which can be viewed as the Banach counterpart of the classical result for tempered distributions (Ludwig, 1966). In Section 4, we use our framework to characterize the sampling functionals (Radon-compatible Diracs) that are in the range of the filtered Radon transform (Theorem 9). In Section 5, we introduce a general definition of a ridge with an arbitrary distributional profile and derive its (filtered) Radon transform. Finally, in Section 6, we apply our formalism to the resolution of a multidimensional supervised-learning problem with a 2nd-order Radon-domain regularization formulated by Parhi and Nowak (2021), the outcome being Theorem 12 on the optimality of ReLU networks.

2. Mathematical Preliminaries

2.1 Notations

We shall consider multidimensional functions f on $\mathbb{R}^d$ that are indexed by the variable $\boldsymbol{x} \in \mathbb{R}^d$. To describe their partial derivatives, we use the multi-index $\boldsymbol{k} = (k_1, \dots, k_d) \in \mathbb{N}^d$ with the notational conventions $\boldsymbol{k}! \triangleq \prod_{i=1}^d k_i!$, $|\boldsymbol{k}| \triangleq k_1 + \cdots + k_d$, $\boldsymbol{x}^{\boldsymbol{k}} \triangleq \prod_{i=1}^d x_i^{k_i}$ for any $\boldsymbol{x} \in \mathbb{R}^d$, and $\partial^{\boldsymbol{k}} f(\boldsymbol{x}) \triangleq \frac{\partial^{|\boldsymbol{k}|} f(x_1,\dots,x_d)}{\partial_{x_1}^{k_1} \cdots \partial_{x_d}^{k_d}}$. This allows us to write the multidimensional Taylor expansion around $\boldsymbol{x}_0$ of an analytical function $f : \mathbb{R}^d \to \mathbb{R}$ explicitly as

$$f(\boldsymbol{x}) = \sum_{n=0}^{\infty} \sum_{|\boldsymbol{k}|=n} \frac{\partial^{\boldsymbol{k}} f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0)^{\boldsymbol{k}}}{\boldsymbol{k}!}, \tag{3}$$

where the internal summation is over all multi-indices $\boldsymbol{k} = (k_1, \dots, k_d)$ such that $k_1 + \cdots + k_d = n$.

Schwartz' space of smooth and rapidly decreasing test functions $\varphi : \mathbb{R}^d \to \mathbb{R}$ equipped with the usual Fréchet-Schwartz topology is denoted by $\mathcal{S}(\mathbb{R}^d)$. Its continuous dual is the space $\mathcal{S}'(\mathbb{R}^d)$ of tempered distributions. In this setting, the Lebesgue spaces $L_p(\mathbb{R}^d)$ for $p \in [1, \infty)$ can be specified as the completion of $\mathcal{S}(\mathbb{R}^d)$ equipped with the L_p-norm $\|\cdot\|_{L_p}$; that is, $L_p(\mathbb{R}^d) = \overline{(\mathcal{S}(\mathbb{R}^d), \|\cdot\|_{L_p})}$. For the end point $p = \infty$, we have that $\overline{(\mathcal{S}(\mathbb{R}^d), \|\cdot\|_{\sup})} = C_0(\mathbb{R}^d)$ with $\|\varphi\|_{\sup} = \sup_{\boldsymbol{x} \in \mathbb{R}^d} |\varphi(\boldsymbol{x})| = \|\varphi\|_{L_\infty}$, where $C_0(\mathbb{R}^d)$ is the space of continuous functions that vanish at infinity. Its continuous dual is the space $\mathcal{M}(\mathbb{R}^d) = \{f \in \mathcal{S}'(\mathbb{R}^d) : \|f\|_{\mathcal{M}} < \infty\}$ of bounded Radon measures with

$$\|f\|_{\mathcal{M}} = \sup_{\varphi \in \mathcal{S}(\mathbb{R}^d) : \|\varphi\|_{L_\infty} \le 1} \langle f, \varphi \rangle.$$

The latter is a superset of $L_1(\mathbb{R}^d)$, which is isometrically embedded in it, meaning that $\|f\|_{L_1} = \|f\|_{\mathcal{M}}$ for all $f \in L_1(\mathbb{R}^d)$.

The Fourier transform of a function $\varphi \in L_1(\mathbb{R}^d)$ is defined as

$$\widehat{\varphi}(\boldsymbol{\omega}) \triangleq \mathcal{F}\{\varphi\}(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi(\boldsymbol{x}) \mathrm{e}^{-\mathrm{j}\langle \boldsymbol{\omega}, \boldsymbol{x} \rangle} \mathrm{d}\boldsymbol{x}. \tag{4}$$

Since the Fourier operator $\mathcal{F}$ continuously maps $\mathcal{S}(\mathbb{R}^d)$ into itself, the transform can be extended by duality to the whole space $\mathcal{S}'(\mathbb{R}^d)$ of tempered distribution. Specifically, $\widehat{f} = \mathcal{F}\{f\} \in \mathcal{S}'(\mathbb{R}^d)$ is the (unique) *generalized Fourier transform* of $f \in \mathcal{S}'(\mathbb{R}^d)$ if and only if $\langle \widehat{f}, \varphi \rangle = \langle f, \widehat{\varphi} \rangle$ for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$, where $\widehat{\varphi} = \mathcal{F}\{\varphi\}$ is the "classical" Fourier transform of φ defined by (4).

2.2 Polynomial Spaces and Related Projectors

The regularization operators (e.g., the Laplacian) that are of interest to us are isotropic and have a growth-restricted null space formed by the polynomials of degree n_0. This space is denoted by $\mathcal{P}_{n_0}$ and is spanned by the monomial/Taylor basis

$$m_{\boldsymbol{k}}(\boldsymbol{x}) \triangleq \frac{\boldsymbol{x}^{\boldsymbol{k}}}{\boldsymbol{k}!} \tag{5}$$

with $|\boldsymbol{k}| \le n_0$. Accordingly, we have that

$$\mathcal{P}_{n_0} = \{p_0 = \sum_{|\boldsymbol{k}| \le n_0} b_{\boldsymbol{k}} m_{\boldsymbol{k}} : \|p_0\|_{\mathcal{P}} < \infty\} \text{ with } \|p_0\|_{\mathcal{P}} \triangleq \|(b_{\boldsymbol{k}})_{|\boldsymbol{k}| \le n_0}\|_2, \tag{6}$$

where we have chosen to equip the space with the ℓ_2-norm of the Taylor coefficients. The important point here is that (6) specifies a finite-dimensional Banach subspace of $\mathcal{S}'(\mathbb{R}^d)$. Its continuous dual $\mathcal{P}'_{n_0}$ is finite-dimensional as well, although it is composed of "abstract" elements that do, in fact, admit infinitely many possible representers in $\mathcal{S}'(\mathbb{R}^d)$. Here, we choose to identify every dual element $p_0^* \in \mathcal{P}'_{n_0}$ concretely as a function in $\mathcal{S}(\mathbb{R}^d)$ by selecting a particular dual basis $\{m^*_{\boldsymbol{k}}\}_{|\boldsymbol{k}| \le n_0}$ such that $\langle m^*_{\boldsymbol{k}}, m_{\boldsymbol{k}'} \rangle = \delta_{\boldsymbol{k}-\boldsymbol{k}'}$ (Kroneker delta). The existence of such a basis is guaranteed because $\mathcal{P}_{n_0}$ is a finite-dimensional subspace of $\mathcal{S}'(\mathbb{R}^d)$ (Rudin, 1991, Theorem 3.5). Our specific choice is

$$m^*_{\boldsymbol{k}} = (-1)^{|\boldsymbol{k}|} \partial^{\boldsymbol{k}} \kappa_{\mathrm{iso}} \in \mathcal{S}(\mathbb{R}^d) \tag{7}$$

with $\boldsymbol{k} \in \mathbb{N}^d$, where κ_{iso} is the isotropic function described below.

Lemma 1 *There exists an entire isotropic function $\kappa_{\mathrm{iso}} \in \mathcal{S}(\mathbb{R}^d)$ with $0 \le \widehat{\kappa}_{\mathrm{iso}}(\boldsymbol{\omega}) \le 1$ and $\widehat{\kappa}_{\mathrm{iso}}(\boldsymbol{\omega}) = 0$ for $\|\boldsymbol{\omega}\| \ge 1$ such that*

$$\int_{\mathbb{R}^d} \frac{\boldsymbol{x}^{\boldsymbol{k}}}{\boldsymbol{k}!} (-1)^{|\boldsymbol{n}|} \partial^{\boldsymbol{n}} \kappa_{\mathrm{iso}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \delta_{\boldsymbol{k}-\boldsymbol{n}} \tag{8}$$

for all $\boldsymbol{k}, \boldsymbol{n} \in \mathbb{N}^d$.

Proof We take $\kappa_{\rm iso} = \mathcal{F}^{-1}\{\widehat{\kappa}_{\rm rad}(\|\cdot\|)\}$, where the radial profile $\widehat{\kappa}_{\rm rad}: \mathbb{R} \to \mathbb{R}$ is such that $\widehat{\kappa}_{\rm rad} \in \mathcal{S}(\mathbb{R})$, $\widehat{\kappa}_{\rm rad}(\omega) = 1$ for $0 \le |\omega| \le R_0 \le \frac{1}{2}$, and $\widehat{\kappa}_{\rm rad}(\omega) = 0$ for $|\omega| \ge 1$. A particular construction with $R_0 = \frac{1}{2}$ is $\widehat{\kappa}_{\rm rad} = {\rm rect} * \varphi$, where $\varphi \in \mathcal{S}(\mathbb{R})$ is a symmetric, non-negative test function (to avoid oscillations) with support$(\varphi) \subseteq [-\frac{1}{2}, \frac{1}{2}]$ and $\int_{\mathbb{R}} \varphi(x){\rm d}x = 1$. Next, we observe that

$$\langle m_{\boldsymbol{k}}, (-1)^{|\boldsymbol{n}|}\partial^{\boldsymbol{n}}\kappa_{\rm iso}\rangle = \langle \partial^{\boldsymbol{n}} m_{\boldsymbol{k}}, \kappa_{\rm iso}\rangle = \begin{cases} \langle m_{\boldsymbol{k}-\boldsymbol{n}}, \kappa_{\rm iso}\rangle, & \boldsymbol{k}-\boldsymbol{n} \ge \boldsymbol{0} \\ \langle 0, \kappa_{\rm iso}\rangle = 0, & \text{otherwise.} \end{cases} \tag{9}$$

We evaluate the duality product for the case $\boldsymbol{m} = (\boldsymbol{k}-\boldsymbol{n}) \ge \boldsymbol{0}$ in the Fourier domain as

$$\begin{aligned} \langle m_{\boldsymbol{m}}, \kappa_{\rm iso}\rangle &= \frac{1}{(2\pi)^d}\langle \mathcal{F}\{m_{\boldsymbol{m}}\}, \widehat{\kappa}_{\rm iso}\rangle \\ &= \langle \frac{{\rm j}^{|\boldsymbol{m}|}}{\boldsymbol{m}!}\delta^{(\boldsymbol{m})}, \widehat{\kappa}_{\rm iso}\rangle = \frac{-{\rm j}^{|\boldsymbol{m}|}}{\boldsymbol{m}!}\partial^{\boldsymbol{m}}\widehat{\kappa}_{\rm iso}(\boldsymbol{0}) = \begin{cases} 1, & \boldsymbol{m} = \boldsymbol{0} \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \tag{10}$$

where we have used the relation $\mathcal{F}\{\boldsymbol{x}^{\boldsymbol{m}}\}(\boldsymbol{\omega}) = (2\pi)^d\ {\rm j}^{|\boldsymbol{m}|}\delta^{(\boldsymbol{m})}(\boldsymbol{\omega})$. Finally, since $\widehat{\kappa}_{\rm iso}$ is compactly supported, its inverse Fourier transform $\kappa_{\rm iso}$ is an entire function of exponential-type (by the Paley-Wiener theorem). This means that the function $\boldsymbol{x} \mapsto \kappa_{\rm iso}(\boldsymbol{x})$ is analytic with a convergent Taylor series of the form (3) for any $\boldsymbol{x}, \boldsymbol{x}_0 \in \mathbb{R}^d$. ■

This allows us to describe the dual space explicitly as

$$\mathcal{P}'_{n_0} = \{p_0^* = \sum_{|\boldsymbol{k}| \le n_0} b^*_{\boldsymbol{k}} m^*_{\boldsymbol{k}} : \|p_0^*\|_{\mathcal{P}'} \stackrel{\triangle}{=} \|(b^*_{\boldsymbol{k}})\|_2 < \infty\} \tag{11}$$

where each elements p_0^* has a unique representation in terms of its coefficients $(b^*_{\boldsymbol{k}})_{|\boldsymbol{k}|\le n_0}$. We also use the dual basis to specify the projection operator ${\rm Proj}_{\mathcal{P}_{n_0}} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{P}_{n_0}$ as

$${\rm Proj}_{\mathcal{P}_{n_0}}\{f\} = \sum_{|\boldsymbol{k}|\le n_0} \langle f, m^*_{\boldsymbol{k}}\rangle\, m_{\boldsymbol{k}}, \tag{12}$$

which is well-defined for any $f \in \mathcal{S}'(\mathbb{R}^d)$ since $m^*_{\boldsymbol{k}} \in \mathcal{S}(\mathbb{R}^d)$. While the form of (12) is generic, it must be emphasized that the resulting projector strongly depends upon our specific choice of dual basis.

2.3 Radon Transform

The Radon transform integrates of a function of $\mathbb{R}^d$ over all hyperplanes of dimension $(d-1)$. These hyperplanes are indexed over $\mathbb{R} \times \mathbb{S}^{d-1}$, where $\mathbb{S}^{d-1} = \{\boldsymbol{\xi} \in \mathbb{R}^d : \|\boldsymbol{\xi}\|_2 = 1\}$ is the unit sphere in $\mathbb{R}^d$. Specifically, the coordinates $\boldsymbol{x}$ of a hyperplane associated with an offset $t \in \mathbb{R}$ and a normal vector $\boldsymbol{\xi} \in \mathbb{S}^{d-1}$ satisfy

$$\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x} = \xi_1 x_1 + \cdots + \xi_d x_d = t. \tag{13}$$

The transform is first described for ordinary (test) functions and then extended to tempered distributions by duality.

2.3.1 Classical Integral Formulation

The Radon transform of the function $f \in L_1(\mathbb{R}^d)$ is defined as

$$\mathrm{R}\{f\}(t, \boldsymbol{\xi}) = \int_{\mathbb{R}^d} \delta(t - \boldsymbol{\xi}^\mathsf{T}\boldsymbol{x}) f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \quad (t, \boldsymbol{\xi}) \in \mathbb{R} \times \mathbb{S}^{d-1}. \tag{14}$$

The adjoint of R is the back-projection operator R^*. Its action on $g \in L_\infty(\mathbb{R} \times \mathbb{S}^{d-1})$ yields the function

$$\mathrm{R}^*\{g\}(\boldsymbol{x}) = \int_{\mathbb{S}^{d-1}} g(\underbrace{\boldsymbol{\xi}^\mathsf{T}\boldsymbol{x}}_{t}, \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}, \quad \boldsymbol{x} \in \mathbb{R}^d, \tag{15}$$

where $\mathrm{d}\boldsymbol{\xi}$ is a surface element on $\mathbb{S}^{d-1}$.

Given the d-dimensional Fourier transform $\hat{f} = \mathcal{F}\{f\} \in C_0(\mathbb{R}^d)$ of $f \in L_1(\mathbb{R}^d)$, one can calculate $\mathrm{R}\{f\}(\cdot, \boldsymbol{\xi}_0)$ for any $\boldsymbol{\xi}_0 \in \mathbb{S}^{d-1}$ with the help of the *Fourier-slice theorem*. The latter is usually stated as

$$\mathcal{F}_{t \to \omega}\big\{\mathrm{R}\{f\}(t, \boldsymbol{\xi}_0)\big\}(\omega) = \int_{\mathbb{R}} \mathrm{R}\{f\}(t, \boldsymbol{\xi}_0) \mathrm{e}^{-\mathrm{j}\omega t} \mathrm{d}t = \hat{f}(\omega \boldsymbol{\xi}_0), \tag{16}$$

which tells us that the restriction of $\hat{f} : \mathbb{R}^d \to \mathbb{C}$ along the ray $\{\boldsymbol{\omega} = \omega \boldsymbol{\xi}_0 : \omega \in \mathbb{R}\}$ is equal to the 1D Fourier transform of $\mathrm{R}\{f\}(\cdot, \boldsymbol{\xi}_0)$.

To describe the functional properties of the Radon transform, one needs the hyper-spherical (or Radon-domain) counterparts of the spaces described in Section 2.1 where the Euclidean indexing with $\boldsymbol{x} \in \mathbb{R}^d$ is replaced by $(t, \boldsymbol{\xi}) \in \mathbb{R} \times \mathbb{S}^{d-1}$. The spherical counterpart of $\mathcal{S}(\mathbb{R}^d)$ is $\mathcal{S}(\mathbb{R} \times \mathbb{S}^{d-1})$. Correspondingly, an element $g \in \mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1})$ is a continuous linear functional on $\mathcal{S}(\mathbb{R} \times \mathbb{S}^{d-1})$ whose action on the test function $\phi(t, \boldsymbol{\xi})$ is represented by the duality product $g : \phi \mapsto \langle g, \phi \rangle_{\mathrm{Rad}}$. When g can be identified as an ordinary function $g : (t, \boldsymbol{\xi}) \mapsto \mathbb{R}$, one has that

$$\langle g, \phi \rangle_{\mathrm{Rad}} = \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} g(t, \boldsymbol{\xi}) \phi(t, \boldsymbol{\xi}) \mathrm{d}t \mathrm{d}\boldsymbol{\xi}, \tag{17}$$

where $\mathrm{d}\boldsymbol{\xi}$ stands for a surface element on $\mathbb{S}^{d-1}$ with $\|\boldsymbol{\xi}\|_2 = 1$. For instance, for $d = 2$, we parameterize $\mathbb{S}^1$ by setting $\boldsymbol{\xi} = (\cos\theta, \sin\theta)$ with $\mathrm{d}\boldsymbol{\xi} = \mathrm{d}\theta$ for $\theta \in [0, 2\pi]$, which then yields

$$\langle g, \phi \rangle_{\mathrm{Rad}} = \int_0^{2\pi} \int_{\mathbb{R}} g(t, \theta) \phi(t, \theta) \mathrm{d}t \mathrm{d}\theta. \tag{18}$$

Such explicit representations are also available in higher dimensions using hyper-spherical polar coordinates. Of special importance to us is the translated and rotated hyper-spherical Dirac distribution $\delta_{(t_0, \boldsymbol{\xi}_0)} = \delta(\cdot - t_0)\delta(\cdot - \boldsymbol{\xi}_0) \in \mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1})$, which is defined as $\langle \delta_{\boldsymbol{z}_0}, \phi \rangle_{\mathrm{Rad}} \triangleq \phi(\boldsymbol{z}_0)$ for all $\phi \in \mathcal{S}(\mathbb{R} \times \mathbb{S}^{d-1})$ and any offset $\boldsymbol{z}_0 = (t_0, \boldsymbol{\xi}_0) \in \mathbb{R} \times \mathbb{S}^{d-1}$. This Dirac impulse, which is separable in the index variables t and $\boldsymbol{\xi}$, is included in the Banach space $\mathcal{M}(\mathbb{R} \times \mathbb{S}^{d-1})$ (hyper-spherical Radon measures) with the property that $\|\delta_{\boldsymbol{z}_0}\|_{\mathcal{M}} = 1$.

The key property for analysis is that the Radon transform is continuous on $\mathcal{S}$ and invertible (see (Ludwig, 1966; Helgason, 2011; Ramm and Katsevich, 2020) for details).

Theorem 2 (Continuity and Invertibility of the Radon Transform on $\mathcal{S}(\mathbb{R}^d)$) *The Radon operator* R *continuously maps* $\mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}\times\mathbb{S}^{d-1})$. *Moreover,* $\mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\mathrm{R} = \mathrm{K}\mathrm{R}^*\mathrm{R} = \mathrm{R}^*\mathrm{R}\mathrm{K} = \mathrm{Id}$ *on* $\mathcal{S}(\mathbb{R}^d)$*, where* $\mathrm{K} = (\mathrm{R}^*\mathrm{R})^{-1} = c_d(-\Delta)^{\frac{d-1}{2}}$ *with* $c_d = \frac{1}{2(2\pi)^{d-1}}$ *is the so-called "filtering" operator, and is* $\mathrm{K}_{\mathrm{rad}}$ *its one-dimensional radial counterpart that acts along the Radon-domain variable* t*. These filtering operators are characterized by their frequency response* $\widehat{K}(\boldsymbol{\omega}) = c_d\|\boldsymbol{\omega}\|^{d-1}$ *and* $\widehat{K}_{\mathrm{rad}}(\omega) = c_d|\omega|^{d-1}$.

The image of $\mathcal{S}(\mathbb{R}^d)$ through the Radon transform is the space $\mathcal{S}_{\mathrm{Rad}} \overset{\triangle}{=} \mathrm{R}\big(\mathcal{S}(\mathbb{R}^d)\big)$: a subset of the space of hyper-spherical test functions $\mathcal{S}(\mathbb{R}\times\mathbb{S}^{d-1})$ that can be characterized explicitly.

Theorem 3 (Gelfand and Shilov (1966); Helgason (2011); Ludwig (1966)) *A hyper-spherical test function* $\phi \in \mathcal{S}(\mathbb{R}\times\mathbb{S}^{d-1})$ *is a valid Radon transform in the sense that* $\phi \in \mathcal{S}_{\mathrm{Rad}} = \{\mathrm{R}\{\varphi\} : \varphi \in \mathcal{S}(\mathbb{R}^d)\}$ *if and only if*

1. *evenness:* $\phi(t,\boldsymbol{\xi}) = \phi(-t,-\boldsymbol{\xi})$.
2. *moment condition: for any* $k \in \mathbb{N}$, $\Phi_k(\boldsymbol{\xi}) = \int_{\mathbb{R}} \phi(t,\boldsymbol{\xi})t^k\mathrm{d}t$ *is a homogeneous polynomial in* $\boldsymbol{\xi} \in \mathbb{S}^{d-1}$.

In particular, a function $\phi = \mathrm{R}\{\varphi\} \in \mathcal{S}_{\mathrm{Rad}}$ must be symmetric and satisfy $\int_{\mathbb{R}} \mathrm{R}\{\varphi\}(t,\boldsymbol{\xi})\mathrm{d}t = \int_{\mathbb{R}^d} \varphi(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$ for all $\boldsymbol{\xi} \in \mathbb{S}^{d-1}$.

While Theorem 2 implies that R is invertible on $\mathcal{S}_{\mathrm{Rad}}$, there is also a stronger version of this property for $\mathcal{S}_{\mathrm{Rad}}$ equipped with the Schwartz-Fréchet topology inherited from $\mathcal{S}(\mathbb{R}\times\mathbb{S}^{d-1})$ (Helgason, 2011, p. 60) (Hertle, 1983).

Theorem 4 *The operator* $\mathrm{R} : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}_{\mathrm{Rad}}$ *is a continuous bijection, with a continuous inverse given by* $\mathrm{R}^{-1} = (\mathrm{R}^*\mathrm{K}_{\mathrm{rad}}) : \mathcal{S}_{\mathrm{Rad}} \to \mathcal{S}(\mathbb{R}^d)$.

The bottom line is that the classical Radon transform is a homeomorphism $\mathcal{S}(\mathbb{R}^d) \to \mathcal{S}_{\mathrm{Rad}}$.

2.3.2 Distributional Extension

To extend the framework to distributions, we proceed by duality along the lines exposed by Ludwig (1966). To that end, we first observe that the inversion formula for the Radon transform on $\mathcal{S}(\mathbb{R}^d)$ and the homeomorphism property in Theorem 4 imply that the inverse Radon transform $\mathrm{R}^{-1} = \mathrm{R}^*\mathrm{K}_{\mathrm{rad}} : \mathcal{S}_{\mathrm{Rad}} \to \mathcal{S}(\mathbb{R}^d)$, as well as the back-projection operator $\mathrm{R}^* : \mathrm{K}_{\mathrm{rad}}(\mathcal{S}_{\mathrm{Rad}}) \to \mathcal{S}(\mathbb{R}^d)$, are endowed with the same property. The idea then is to identify the adjoint of these operators which act on the dual spaces of distributions $\mathcal{S}'(\mathbb{R}^d) = \big(\mathcal{S}(\mathbb{R}^d)\big)'$, $\mathcal{S}'_{\mathrm{Rad}} = \big(\mathcal{S}_{\mathrm{Rad}}\big)'$, and $\mathcal{S}'(\mathbb{R}\times\mathbb{S}^{d-1}) = \big(\mathcal{S}(\mathbb{R}\times\mathbb{S}^{d-1})\big)'$ in the unrestricted scenario. The additional ingredient is the symmetry of the radial filtering operator $\mathrm{K}_{\mathrm{rad}}$, which translates into $\mathrm{K}_{\mathrm{rad}} = \mathrm{K}^*_{\mathrm{rad}}$.

Definition 5 (Generalized Radon transform, filtering and back-projection)

1. *The distributional Radon transform*

$$\mathrm{R} : \mathcal{S}'(\mathbb{R}^d) \to \big(\mathrm{K}_{\mathrm{rad}}(\mathcal{S}_{\mathrm{Rad}})\big)'$$

is defined as the dual map of the homeomorphism $\mathrm{R}^* : \mathrm{K}_{\mathrm{rad}}(\mathcal{S}_{\mathrm{Rad}}) \to \mathcal{S}(\mathbb{R}^d)$*, where* R^* *is well-defined by* (15) *since* $\mathrm{K}_{\mathrm{rad}}(\mathcal{S}_{\mathrm{Rad}}) \subset L_\infty(\mathbb{R}^d)$.

2. *The distributional filtered projection*

$$\mathrm{K}_{\mathrm{rad}}\mathrm{R} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'_{\mathrm{Rad}}$$

is defined as the dual map of the homeomorphism $\mathrm{R}^*\mathrm{K}_{\mathrm{rad}} : \mathcal{S}_{\mathrm{Rad}} \to \mathcal{S}(\mathbb{R}^d)$.

3. *The restricted distributional back-projection*

$$\mathrm{R}^* : \mathcal{S}'_{\mathrm{Rad}} \to \mathcal{S}'(\mathbb{R}^d)$$

is defined as the dual map of the homeomorphism $\mathrm{R} : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}_{\mathrm{Rad}}$*, where* R *is well-defined by* (14) *since* $\mathcal{S}(\mathbb{R}^d) \subset L_1(\mathbb{R}^d)$.

4. *The unrestricted distributional back-projection*

$$\mathrm{R}^* : \mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1}) \to \mathcal{S}'(\mathbb{R}^d)$$

is defined as the dual map of the continuous operator $\mathrm{R} : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R} \times \mathbb{S}^{d-1})$*, which takes advantage of the property that* $\mathcal{S}_{\mathrm{Rad}}$ *is continuously embedded in* $\mathcal{S}(\mathbb{R} \times \mathbb{S}^{d-1})$.

Based on these definitions, one obtains the classical result on the invertibility of the (filtered) Radon transform on $\mathcal{S}'(\mathbb{R}^d)$ (Ludwig, 1966), which is the dual of Theorem 4.

Theorem 6 (Invertibility of the Radon Transform on $\mathcal{S}'(\mathbb{R}^d)$) *It holds that* $\mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\mathrm{R} = \mathrm{Id}$ *on* $\mathcal{S}'(\mathbb{R}^d)$. *Moreover, the filtered Radon transform* $\mathrm{K}_{\mathrm{rad}}\mathrm{R} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'_{\mathrm{Rad}}$ *is bicontinuous and one-to-one, with its continuous inverse* $(\mathrm{K}_{\mathrm{rad}}\mathrm{R})^{-1}$ *being the back-projection operation* $\mathrm{R}^* : \mathcal{S}'_{\mathrm{Rad}} \to \mathcal{S}'(\mathbb{R}^d)$.

While Theorem 6 ensures that the filtered Radon transform is well-defined in the distributional sense and invertible on $\mathcal{S}'_{\mathrm{Rad}}$, it is an abstract characterization because the range of the operator is an equivalence class of distributions. In fact, one can show that $\mathcal{S}'_{\mathrm{Rad}}$ does not have a closed complement in $\mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1})$, which means that $\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{f\} \in \mathcal{S}'_{\mathrm{Rad}}$ cannot, in general, be identified as a unique element in $\mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1})$. The situation is more favorable for the back-projection operator R^*, which admits the continuous extension $\mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1}) \to \mathcal{S}'(\mathbb{R}^d)$ (see fourth item (unrestricted operator) in Definition 6). The latter then yields a *concrete* characterization of the (generalized) back-projection of any hyper-spherical distribution. Specifically, $f = \mathrm{R}^*\{g\} \in \mathcal{S}'(\mathbb{R}^d)$ is the back-projection of $g \in \mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1})$ if

$$\forall \varphi \in \mathcal{S}(\mathbb{R}^d) : \quad \langle \mathrm{R}^*\{g\}, \varphi \rangle = \langle g, \mathrm{R}\{\varphi\} \rangle_{\mathrm{Rad}}. \tag{19}$$

Likewise, we can specify the null space of this unrestricted operator as

$$\mathcal{N}_{\mathrm{R}^*} = \{g \in \mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1}) : \mathrm{R}^*\{g\} = 0 \;\Leftrightarrow\; \langle g, \phi \rangle_{\mathrm{Rad}} = 0, \forall \phi \in \mathcal{S}_{\mathrm{Rad}}\}. \tag{20}$$

It is important to appreciate that this null space is huge: In addition to the odd hyper-spherical distributions, it includes a sizeable subset of the even distributions because of the moment condition in Theorem 3. Since $\mathcal{N}_{\mathrm{R}^*} = \mathrm{null}(\mathrm{R}^*)$ is a closed subspace of $\mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1})$, we can now identify $\mathcal{S}'_{\mathrm{Rad}}$ as the abstract quotient space $\mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1})/\mathcal{N}_{\mathrm{R}^*}$. In other words, if, for a given $f \in \mathcal{S}'(\mathbb{R}^d)$, we find a hyper-spherical distribution $g_0 \in \mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1})$ such that

$$\forall \phi \in \mathcal{S}_{\mathrm{Rad}}: \quad \langle g_0, \phi \rangle_{\mathrm{Rad}} = \langle f, \mathrm{R}^* \mathrm{K}_{\mathrm{rad}}\{\phi\} \rangle, \tag{21}$$

then, strictly speaking, $\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{f\} \in \mathcal{S}'_{\mathrm{Rad}}$ is the equivalence class (or coset) given by

$$\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{f\} = [g_0] = \{g_0 + h : h \in \mathcal{N}_{\mathrm{R}^*}\}. \tag{22}$$

The members of $[g_0]$ are interchangeable—we refer to them as "formal" filtered projections of f to remind us of this lack of unicity.

To illustrate this ambiguity, we consider the Dirac ridge $\delta(\boldsymbol{\xi}_0^{\mathsf{T}}\boldsymbol{x} - t_0) \in \mathcal{S}'(\mathbb{R}^d)$ and refer to Definition (14) of the Radon transform to deduce that, for all $\phi = \mathrm{R}\{\varphi\} \in \mathcal{S}_{\mathrm{Rad}}$ with $\varphi \in \mathcal{S}(\mathbb{R}^d)$,

$$\begin{aligned}\langle \delta(\boldsymbol{\xi}_0^{\mathsf{T}} \cdot - t_0), \mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\{\phi\}\rangle &= \langle \delta(\boldsymbol{\xi}_0^{\mathsf{T}} \cdot - t_0), \overbrace{\mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\mathrm{R}}^{\mathrm{Id}}\{\varphi\}\rangle \\ &= \int_{\mathbb{R}^d} \delta(\boldsymbol{\xi}_0^{\mathsf{T}}\boldsymbol{x} - t_0)\varphi(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \mathrm{R}\{\varphi\}(-\boldsymbol{z}_0) = \langle \delta_{-\boldsymbol{z}_0}, \phi\rangle_{\mathrm{Rad}},\end{aligned} \tag{23}$$

where $\boldsymbol{z}_0 = (t_0, \boldsymbol{\xi}_0)$, which shows that the Dirac impulse $\delta_{-\boldsymbol{z}_0}$ is a formal filtered projection of $\delta(\boldsymbol{\xi}_0^{\mathsf{T}}\boldsymbol{x} - t_0)$. Moreover, since $\delta(\boldsymbol{\xi}_0^{\mathsf{T}}\boldsymbol{x} - t_0) = \delta(-\boldsymbol{\xi}_0^{\mathsf{T}}\boldsymbol{x} + t_0)$, the same holds true for $\delta_{\boldsymbol{z}_0}$ as well as for $\frac{1}{2}\big(\delta_{\boldsymbol{z}_0} + \delta_{-\boldsymbol{z}_0}\big)$. In fact, there is an infinity of potential "formal" solutions, which is sumarized as

$$\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{\delta(\boldsymbol{\xi}_0^{\mathsf{T}} \cdot - t_0)\} = [\delta_{(t_0, \boldsymbol{\xi}_0)}] \in \mathcal{S}'_{\mathrm{Rad}}. \tag{24}$$

The distributional extension of the Radon transform inherits most of the properties of the "classical" operator defined in (14). Of special relevance to us is the quasi-commutativity of R with convolution, also known as the *intertwining property*. Specifically, let $h, f \in \mathcal{S}'(\mathbb{R}^d)$ be two distributions whose convolution $h * f$ is well defined in $\mathcal{S}'(\mathbb{R}^d)$. Then,

$$\mathrm{R}\{h * f\} = \mathrm{R}\{h\} \circledast \mathrm{R}\{f\} \tag{25}$$

where the symbol "$\circledast$" denotes a convolution along the radial variable t; that is, $(u \circledast g)(t, \boldsymbol{\xi}) = \langle u(\cdot, \boldsymbol{\xi}), g(t - \cdot, \boldsymbol{\xi})\rangle$. In particular, when $h = \mathrm{L}\{\delta\}$ is the (isotropic) impulse response of a linear shift-invariant (LSI) operator whose frequency response $\widehat{L}(\boldsymbol{\omega}) = \widehat{L}_{\mathrm{rad}}(\|\boldsymbol{\omega}\|)$ is purely radial, we get that

$$\mathrm{R}\{h * f\} = \mathrm{R}\mathrm{L}\{f\} = \mathrm{L}_{\mathrm{rad}}\mathrm{R}\{f\}, \tag{26}$$

where $\mathrm{L}_{\mathrm{rad}}$ is the $\circledast$-convolution operator whose 1D frequency response is $\widehat{L}_{\mathrm{rad}}(\omega)$. Likewise, by duality, for $g \in \mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1})$ we have that

$$\mathrm{L}\mathrm{R}^*\{g\} = \mathrm{R}^*\mathrm{L}_{\mathrm{rad}}\{g\} \tag{27}$$

under the implicit assumption that $\mathrm{L}\{\mathrm{R}^*g\}$ and $\mathrm{L}_{\mathrm{rad}}\{g\}$ are well-defined distributions. By taking inspiration from Theorem 2, we may then use Relations (26) and (27) for $\mathrm{L} = \mathrm{K} =$

$(\mathrm{R}^*\mathrm{R})^{-1}$ to show that $\mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{f\} = \mathrm{R}^*\mathrm{R}\mathrm{K}\{f\} = \mathrm{K}\mathrm{R}^*\mathrm{R}\{f\} = f$ for a broad class of distributions. While the first form is valid for all $f \in \mathcal{S}'(\mathbb{R}^d)$ (see Theorem 6), there is a slight restriction with the second (resp., third), which requires that $\mathrm{K}\{f\}$ (resp., $\mathrm{K}\{g\}$ with $g = \mathrm{R}^*\mathrm{R}\{f\} \in \mathcal{S}'(\mathbb{R}^d)$) be well-defined in $\mathcal{S}'(\mathbb{R}^d)$. While the latter condition is always met when d is odd, it can fail in even dimensions for distributions (e.g., polynomials) whose Fourier transform is singular at the origin[1].

The Fourier-slice theorem expressed by (16) also yields a unique (Fourier-based) characterization of $\mathrm{R}\{f\}$, which remains valid for $f \in \mathcal{S}'(\mathbb{R}^d)$ (Ramm and Katsevich, 2020). It is especially helpful when the underlying function or distribution is isotropic. An isotropic function $\rho_{\mathrm{iso}} : \mathbb{R}^d \to \mathbb{R}$ is characterized by its radial profile $\rho : \mathbb{R}_{\geq 0} \to \mathbb{R}$, with $\rho_{\mathrm{iso}}(\boldsymbol{x}) = \rho(\|\boldsymbol{x}\|)$. The frequency-domain counterpart of this characterization is $\widehat{\rho}_{\mathrm{iso}}(\boldsymbol{\omega}) = \widehat{\rho}_{\mathrm{rad}}(\|\boldsymbol{\omega}\|)$ where the radial frequency profile can be computed as

$$\widehat{\rho}_{\mathrm{rad}}(\omega) = \frac{(2\pi)^{d/2}}{|\omega|^{d/2-1}} \int_0^{+\infty} \rho(t) t^{d/2-1} J_{d/2-1}(\omega t) t \mathrm{d}t, \tag{28}$$

where J_ν is the Bessel function of the first kind of order ν.

Proposition 7 (Radon Transform of Isotropic Distributions) *Let ρ_{iso} be an isotropic distribution whose radial frequency profile is $\widehat{\rho}_{\mathrm{rad}}(\omega)$. Then,*

$$\mathrm{R}\{\rho_{\mathrm{iso}}(\cdot - \boldsymbol{x}_0)\}(t, \boldsymbol{\xi}) = \rho_{\mathrm{rad}}(t - \boldsymbol{\xi}^\mathsf{T}\boldsymbol{x}_0) \tag{29}$$

$$\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{\rho_{\mathrm{iso}}(\cdot - \boldsymbol{x}_0)\}(t, \boldsymbol{\xi}\} = \tilde{\rho}_{\mathrm{rad}}(t - \boldsymbol{\xi}^\mathsf{T}\boldsymbol{x}_0) \tag{30}$$

$$\mathrm{R}\{\partial^{\boldsymbol{m}}\rho_{\mathrm{iso}}\}(t, \boldsymbol{\xi}) = \boldsymbol{\xi}^{\boldsymbol{m}}\mathrm{D}^{|\boldsymbol{m}|}\{\rho_{\mathrm{rad}}\}(t) \tag{31}$$

with $\rho_{\mathrm{rad}}(t) = \mathcal{F}^{-1}\{\widehat{\rho}_{\mathrm{rad}}(\omega)\}(t)$ *and* $\tilde{\rho}_{\mathrm{rad}}(t) = \frac{1}{2(2\pi)^{d-1}}\mathcal{F}^{-1}\{|\omega|^{d-1}\widehat{\rho}_{\mathrm{rad}}(\omega)\}(t)$.

Proof These identities are direct consequences of the Fourier-slice theorem. For instance, by setting $\boldsymbol{\omega} = \omega\boldsymbol{\xi}$ in the Fourier transform of $\partial^{\boldsymbol{m}}\rho_{\mathrm{iso}}$, we get that

$$\widehat{\partial^{\boldsymbol{m}}\rho_{\mathrm{iso}}}(\omega\boldsymbol{\xi}) = (\mathrm{j}\omega\boldsymbol{\xi})^{\boldsymbol{m}}\widehat{\rho}_{\mathrm{rad}}(\omega) = \boldsymbol{\xi}^{\boldsymbol{m}}(\mathrm{j}\omega)^{|\boldsymbol{m}|}\widehat{\rho}_{\mathrm{rad}}(\omega) \tag{32}$$

which, upon taking the inverse 1D Fourier transform, yields (31). ■

Let us note that both ρ_{rad} and $\tilde{\rho}_{\mathrm{rad}}$, as inverse Fourier transform of a real-valued function, are symmetric, which is consistent with the symmetry of the Radon transform and its filtered version.

3. Radon-Compatible Banach Spaces

Our investigation of functional-optimization problems with Radon-domain regularization requires some Banach counterparts of Theorems 4 and 6, preferably such that both $\mathrm{R}\{f\}$ and

1. For $d = 2n$ even, $\widehat{K}(\boldsymbol{\omega}) \propto \|\boldsymbol{\omega}\|^{2n-1}$ which is C^∞ everywhere except at the origin, where it is only C^{2n-2}, meaning that K can only properly handle (and annihilate) polynomials up to degree $(2n-2)$.

$\mathrm{K}_{\mathrm{Rad}}\mathrm{R}\{f\}$ have concrete representations as hyper-spherical functions or measures. In particular, we are interested in identifying specific Radon-domain Banach spaces—for instance, an appropriate subspace of hyper-spherical measures—over which the back-projection operator R^* is guaranteed to be invertible.

To that end, we pick a "parent" hyper-spherical Banach space $\mathcal{X} = (\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ such that $\mathcal{S}(\mathbb{R}\times\mathbb{S}^{d-1}) \overset{\mathrm{d.}}{\hookrightarrow} \mathcal{X} \overset{\mathrm{d.}}{\hookrightarrow} \mathcal{S}'(\mathbb{R}\times\mathbb{S}^{d-1})$. This dense embedding hypothesis has several implications.

1. The Banach space $\mathcal{X}$ is the completion of $\mathcal{S}(\mathbb{R}\times\mathbb{S}^{d-1})$ in the $\|\cdot\|_{\mathcal{X}}$ norm; i.e.,
$$\mathcal{X} = \overline{\big(\mathcal{S}(\mathbb{R}\times\mathbb{S}^{d-1}), \|\cdot\|_{\mathcal{X}}\big)}. \tag{33}$$

2. The dual space $\mathcal{X}' \hookrightarrow \mathcal{S}'(\mathbb{R}\times\mathbb{S}^{d-1})$ is equipped with the norm
$$\|g\|_{\mathcal{X}'} = \sup_{\phi\in\mathcal{X}:\, \|\phi\|_{\mathcal{X}}\le 1} \langle g,\phi\rangle = \sup_{\phi\in\mathcal{S}(\mathbb{R}\times\mathbb{S}^{d-1}):\, \|\phi\|_{\mathcal{X}}\le 1} \langle g,\phi\rangle, \tag{34}$$
where the restriction of $\phi\in\mathcal{S}(\mathbb{R}\times\mathbb{S}^{d-1})$ on the rightmost side of (34) is justified by the denseness of $\mathcal{S}(\mathbb{R}\times\mathbb{S}^{d-1})$ in $\mathcal{X}$.

3. The definition of $\|g\|_{\mathcal{X}'}$ found in the rightmost side of (34) is valid for any distribution $g\in\mathcal{S}'(\mathbb{R}\times\mathbb{S}^{d-1})$ with $\|g\|_{\mathcal{X}'}=\infty$ for $g\notin\mathcal{X}'$. Accordingly, we can specify the topological dual of $\mathcal{X}'$ as
$$\mathcal{X}' = \big\{g\in\mathcal{S}'(\mathbb{R}\times\mathbb{S}^{d-1}) : \|g\|_{\mathcal{X}'}<\infty\big\}. \tag{35}$$

Prototypical examples where those properties are met are $(\mathcal{X},\mathcal{X}') = \big(L_p(\mathbb{R}\times\mathbb{S}^{d-1}), L_q(\mathbb{R}\times\mathbb{S}^{d-1})\big)$ with $p\in[1,\infty)$ and $q=p/(p-1)$ (conjugate exponent), as well as $(\mathcal{X},\mathcal{X}') = \big(C_0(\mathbb{R}\times\mathbb{S}^{d-1}), \mathcal{M}(\mathbb{R}\times\mathbb{S}^{d-1})\big)$ for $p=\infty$.

Likewise, by considering the dual pair $(\mathcal{S}_{\mathrm{Rad}}, \mathcal{S}'_{\mathrm{Rad}})$, we specify our Radon-compatible Banach subspaces
$$\mathcal{X}_{\mathrm{Rad}} = \overline{(\mathcal{S}_{\mathrm{Rad}}, \|\cdot\|_{\mathcal{X}})} \tag{36}$$
$$\mathcal{X}'_{\mathrm{Rad}} = \big(\mathcal{X}_{\mathrm{Rad}}\big)' = \big\{g\in\mathcal{S}'_{\mathrm{Rad}} : \|g\|_{\mathcal{X}'_{\mathrm{Rad}}}<\infty\big\}, \tag{37}$$
where the underlying dual norms have a definition that is analogous to (34) with $\mathcal{S}(\mathbb{R}\times\mathbb{S}^{d-1})$ and $\mathcal{X}$ being instanciated by $\mathcal{S}_{\mathrm{Rad}}$ and $\mathcal{X}_{\mathrm{Rad}}$. We now show that R^* (resp., $\mathrm{K}_{\mathrm{rad}}\mathrm{R}^*$) is invertible on $\mathcal{X}'_{\mathrm{Rad}}$ (resp., on $\mathcal{X}_{\mathrm{Rad}}$), which is the main theoretical contribution of this work.

Since $\mathrm{R}: \mathcal{S}(\mathbb{R}^d)\to\mathcal{S}_{\mathrm{Rad}}$ is a homeomorphism and the $\|\cdot\|_{\mathcal{X}}$-norm is continuous in the Fréchet topology of $\mathcal{S}_{\mathrm{Rad}}$ $\big($resp., of $\mathcal{S}(\mathbb{R}\times\mathbb{S}^{d-1})\big)$, we can consider the normed space $(\mathcal{S}(\mathbb{R}^d), \|\cdot\|_{\mathcal{Y}})$ with $\|\varphi\|_{\mathcal{Y}} \overset{\Delta}{=} \|\mathrm{R}\{\varphi\}\|_{\mathcal{X}}$ and identify the operator $\mathrm{R}: (\mathcal{S}(\mathbb{R}^d), \|\cdot\|_{\mathcal{Y}}) \to (\mathcal{S}_{\mathrm{Rad}}, \|\cdot\|_{\mathcal{X}})$ as an isometry. We then invoke the BLT theorem (Reed and Simon, 1980) to specify the unique extension $\mathrm{R}: \mathcal{Y}\to\overline{(\mathcal{S}_{\mathrm{Rad}}, \|\cdot\|_{\mathcal{X}})} = \mathcal{X}_{\mathrm{Rad}}$, where $\mathcal{Y}$ is a Banach space isometric to $\mathcal{X}_{\mathrm{Rad}}$. More precisely, we have that
$$\mathcal{Y} = \overline{(\mathcal{S}(\mathbb{R}^d), \|\cdot\|_{\mathcal{Y}})} = \{\mathrm{R}^{-1}\{g\} : g\in\mathcal{X}_{\mathrm{Rad}}\} \tag{38}$$
with $\mathrm{R}^{-1} = \mathrm{R}^*\mathrm{K}_{\mathrm{rad}}$ on $\mathcal{S}_{\mathrm{Rad}}$ and, by extension, on $\mathcal{X}_{\mathrm{Rad}}$. This then leads to the following characterization.

Theorem 8 (Radon-compatible Banach spaces) *Let* $(\mathcal{X}_{\mathrm{Rad}}, \mathcal{X}'_{\mathrm{Rad}})$ *be a dual pair of hyper-spherical Banach spaces induced by a norm* $\|\cdot\|_{\mathcal{X}}$ *and specified by* (36) *and* (37). *Then, the following properties hold.*

1. *The back-projection* $\mathrm{R}^*: \mathcal{X}'_{\mathrm{Rad}} \to \mathcal{S}'(\mathbb{R}^d)$ *is injective and* $\mathrm{K_{rad}RR^*} = \mathrm{Id}$ *on* $\mathcal{X}'_{\mathrm{Rad}}$.
2. *The filtered back-projection* $\mathrm{R^*K_{rad}}: \mathcal{X}_{\mathrm{Rad}} \to \mathcal{Y}$ *is an isometric bijection, with* $\mathrm{RR^*K_{rad}} = \mathrm{Id}$ *on* $\mathcal{X}_{\mathrm{Rad}}$.
3. *The corresponding "range" spaces* $\mathcal{Y} = \mathrm{R^*K_{rad}}(\mathcal{X}_{\mathrm{Rad}})$ *and* $\mathcal{Y}' = \mathrm{R}^*(\mathcal{X}'_{\mathrm{Rad}})$ *form a dual Banach pair that is isomorphic to* $(\mathcal{X}_{\mathrm{Rad}}, \mathcal{X}'_{\mathrm{Rad}})$.

Moreover, if there exists a complementary Banach space $\mathcal{X}^{\mathrm{c}}_{\mathrm{Rad}}$ *such that* $\mathcal{X} = \mathcal{X}_{\mathrm{Rad}} \oplus \mathcal{X}^{\mathrm{c}}_{\mathrm{Rad}}$, *then additional properties hold.*

1'. *The dual space is decomposable as* $\mathcal{X}' = \mathcal{X}'_{\mathrm{Rad}} \oplus (\mathcal{X}^{\mathrm{c}}_{\mathrm{Rad}})'$.

2'. *The dual complement* $(\mathcal{X}^{\mathrm{c}}_{\mathrm{Rad}})'$ *is the null space of* $\mathrm{R}^*: \mathcal{X}' \to \mathcal{Y}'$.

3'. *The complement space* $\mathcal{X}^{\mathrm{c}}_{\mathrm{Rad}}$ *is the null space of* $\mathrm{R^*K_{rad}}: \mathcal{X} \to \mathcal{Y}$.

4'. *The operators* $\mathrm{P_{Rad}} = \mathrm{RR^*K_{rad}}: \mathcal{X} \to \mathcal{X}'_{\mathrm{Rad}}$ *and* $\mathrm{P^*_{Rad}} = \mathrm{K_{rad}RR^*}: \mathcal{X}' \to \mathcal{X}_{\mathrm{Rad}}$ *form an adjoint pair of continuous projectors with* $\mathrm{P_{Rad}}\big(\mathcal{X}\big) = \mathcal{X}_{\mathrm{Rad}}$ *and* $\mathrm{P^*_{Rad}}\big(\mathcal{X}'\big) = \mathcal{X}'_{\mathrm{Rad}}$.

Proof

1. Since $\mathcal{S}_{\mathrm{Rad}}$ is densely embedded in $\mathcal{X}_{\mathrm{Rad}}$, we have that $\mathcal{X}'_{\mathrm{Rad}} \hookrightarrow \mathcal{S}'_{\mathrm{Rad}}$ by duality. Let $\mathcal{V} = \mathrm{R}^*(\mathcal{X}'_{\mathrm{Rad}}) \subset \mathcal{S}'(\mathbb{R}^d)$ be the range of the restricted operator $\mathrm{R}^*|_{\mathcal{X}'_{\mathrm{Rad}}}$. Clearly, this restriction is a homeomorphism from $\mathcal{X}'_{\mathrm{Rad}} \to \mathcal{V}$, where we equip $\mathcal{V}$ with the norm $\|f\|_{\mathcal{V}} \triangleq \|\mathrm{R}^{*-1} f\|_{\mathcal{X}'_{\mathrm{Rad}}}$ so that $\mathrm{R}^*: \mathcal{X}'_{\mathrm{Rad}} \to \mathcal{V}$ is an isometric isomorphism. This shows that $\mathrm{R}^*: \mathcal{X}'_{\mathrm{Rad}} \to \mathcal{S}'(\mathbb{R}^d)$ is an injection. Moreover, by Theorem 6, the inverse of $\mathrm{R}^*|_{\mathcal{X}'_{\mathrm{Rad}}}$ on $\mathcal{V}$ is given by the restriction of $\mathrm{K_{rad}R}$ to $\mathcal{V}$, denoted $\mathrm{K_{rad}R}|_{\mathcal{V}}$. Therefore, $\mathrm{K_{rad}RR^*} = \mathrm{Id}$ on $\mathcal{X}'_{\mathrm{Rad}}$.

2. The continuity of $\mathrm{K_{rad}R}: \mathcal{V} \to \mathcal{X}'_{\mathrm{Rad}}$ from Item 1, together with the isometric embedding of a Banach space in its bidual, implies the continuity of $\mathrm{R^*K_{rad}}: \mathcal{X}_{\mathrm{Rad}} \to \mathcal{V}'$ (isometry). The condition $\mathrm{R^*K_{rad}}\{\phi\} = 0$ can then be restated as $\forall v \in \mathcal{V}: \langle v, \mathrm{R^*K_{rad}}\{\phi\}\rangle = \langle \mathrm{K_{rad}R}\{v\}, \phi\rangle_{\mathrm{Rad}} = 0$. Since $\mathcal{X}'_{\mathrm{Rad}} = \mathrm{K_{rad}R}\big(\mathcal{V}\big)$, this is equivalent to say that $\langle g, \phi\rangle_{\mathrm{Rad}} = 0$ for all $g \in \mathcal{X}'_{\mathrm{Rad}}$, which leads to $\phi = 0$ and proves that $\mathrm{R^*K_{rad}}$ is injective on $\mathcal{X}_{\mathrm{Rad}}$. From the characterization given by (38), we readily deduce that $\mathcal{Y} = \mathrm{R^*K_{rad}}(\mathcal{X}_{\mathrm{Rad}}) \subseteq \mathcal{V}'$. Since $\mathrm{R^*K_{rad}}$ is continuous and injective on $\mathcal{X}_{\mathrm{Rad}}$, by the bounded inverse theorem (Rudin, 1991), it has a continuous inverse that maps $\mathcal{Y} \to \mathcal{X}_{\mathrm{Rad}}$. Putting everything together, we get that $\mathrm{RK_{rad}}: \mathcal{X}_{\mathrm{rad}} \to \mathcal{Y}$ is an isometric bijection with inverse given by $(\mathrm{R^*K_{rad}})^{-1} = \mathrm{R}: \mathcal{Y} \to \mathcal{X}_{\mathrm{Rad}}$, which also yields that $\mathrm{RR^*K_{rad}} = \mathrm{Id}$ on $\mathcal{X}_{\mathrm{Rad}}$.

3. From Item 2, we have that $\mathrm{R^*K_{rad}}: \mathcal{X}_{\mathrm{Rad}} \to \mathcal{Y}$ and $\mathrm{R}: \mathcal{Y} \to \mathcal{X}_{\mathrm{Rad}}$ are homeomorphisms. By duality, this implies that

$$
\begin{aligned}
\mathrm{K_{rad}R}: \mathcal{Y}' &\to \mathcal{X}'_{\mathrm{Rad}}\\
\mathrm{R}^*: \mathcal{X}'_{\mathrm{Rad}} &\to \mathcal{Y}'
\end{aligned}
$$

are homeomorphisms and, in particular, bijective isometries with $\mathcal{Y}' = \mathrm{R}^*(\mathcal{X}'_{\mathrm{Rad}})$. Moreover, from the proof of Item 1, we know that $\mathcal{X}'_{\mathrm{Rad}}$ is isometrically isomorphic to $\mathcal{V} = \mathrm{R}^*(\mathcal{X}'_{\mathrm{Rad}})$, which allows us to identify $\mathcal{Y}' = \mathcal{V}$.

1'. This follows from the generic property that $(\mathcal{X}_1 \oplus \mathcal{X}_2)' = \mathcal{X}'_1 \oplus \mathcal{X}'_2$, where the Banach spaces $\mathcal{X}_1$ and $\mathcal{X}_2$ can be arbitrary.

2'. Item 1' tells us that $\mathcal{X}'_{\mathrm{Rad}}, (\mathcal{X}^{\mathrm{c}}_{\mathrm{Rad}})' \subseteq \mathcal{X}'$ with the embedding being continuous. Moreover, it implies that $\mathcal{X}_{\mathrm{Rad}}$ is the annihilator of $\mathcal{X}'_{\mathrm{Rad}}$ in $\mathcal{X}'$, with

$$\begin{aligned}(\mathcal{X}^{\mathrm{c}}_{\mathrm{Rad}})' &= \{g \in \mathcal{X}' : \langle g, \phi\rangle_{\mathrm{Rad}} = 0, \forall \phi \in \mathcal{X}_{\mathrm{Rad}}\} \\ &= \{g \in \mathcal{X}' : \langle g, \phi\rangle_{\mathrm{Rad}} = 0, \forall \phi \in \mathcal{S}_{\mathrm{Rad}}\}, \end{aligned} \tag{39}$$

where the substitution of $\mathcal{X}_{\mathrm{Rad}}$ with $\mathcal{S}_{\mathrm{Rad}}$ in (39) is legitimate because $\mathcal{S}_{\mathrm{Rad}}$ is a dense subspace of $\mathcal{X}_{\mathrm{Rad}}$. Since $\mathcal{X}' \subset \mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1})$, this shows that $(\mathcal{X}^{\mathrm{c}}_{\mathrm{Rad}})' \subseteq \mathcal{N}_{\mathrm{R}^*}$.

3'. This is the dual of Item 1'.

4'. The null-space property implies that $\mathrm{P}^*_{\mathrm{Rad}}\big((\mathcal{X}^{\mathrm{c}}_{\mathrm{Rad}})'\big) = \{0\}$ and $\mathrm{P}_{\mathrm{Rad}}\big(\mathcal{X}^{\mathrm{c}}_{\mathrm{Rad}}\big) = \{0\}$. The final element is the identity in Item 1 (resp., Item 2), which ensures that $\mathrm{P}^*_{\mathrm{Rad}} : \mathcal{X}' \to \mathcal{X}'_{\mathrm{Rad}}$ (resp., $\mathrm{P}_{\mathrm{Rad}} : \mathcal{X} \to \mathcal{X}_{\mathrm{Rad}}$) is the canonical projector on $\mathcal{X}'_{\mathrm{Rad}}$ (resp., on $\mathcal{X}_{\mathrm{Rad}}$). The existence (unicity) and continuity of the latter is guaranteed for any pair of complemented Banach spaces (see Appendix A).

■

The dual direct-sum decomposition in Theorem 8 has the following corollary: The Banach complement $\mathcal{X}^{\mathrm{c}}_{\mathrm{Rad}}$ (resp., $\mathcal{X}_{\mathrm{Rad}}$) is the annihilator of $\mathcal{X}'_{\mathrm{Rad}}$ in $\mathcal{X}$ (resp., of $(\mathcal{X}^{\mathrm{c}}_{\mathrm{Rad}})'$ in $\mathcal{X}'$), and vice versa. In particular, $(\mathcal{X}^{\mathrm{c}}_{\mathrm{Rad}})' \hookrightarrow \mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1})$ is specified by (39) (as a set), which clearly shows that $(\mathcal{X}^{\mathrm{c}}_{\mathrm{Rad}})' \subseteq \mathcal{N}_{\mathrm{R}^*}$ (see (20)).

The existence of the projection operator $\mathrm{P}^*_{\mathrm{Rad}}$ in the second part of Theorem 8 is especially useful when the "abstract" elements of the dual space $\mathcal{X}'$ can be identified as bona-fide hyper-spherical functions or measures. It then enables us to convert any "formal" filtered projection $\tilde{g} = \mathrm{K}_{\mathrm{rad}}\mathrm{R}\{f\} \in \mathcal{X}' \hookrightarrow \mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1})$ (see (21) in Definition 5) into a concrete representer $\mathrm{P}^*_{\mathrm{Rad}}\{\tilde{g}\} \in \mathcal{X}'_{\mathrm{Rad}}$, which has a unique, unambiguous interpretation.

Since the members of $\mathcal{S}_{\mathrm{Rad}}$ must be even (see Theorem 3), we readily deduce that $\mathcal{X}_{\mathrm{Rad}} \subseteq \mathcal{X}_{\mathrm{even}} = \overline{\big(\mathcal{S}_{\mathrm{even}}(\mathbb{R} \times \mathbb{S}^{d-1}), \|\cdot\|_{\mathcal{X}}\big)}$. In particular, when the inclusion is a set equality and when $\mathcal{X} = \mathcal{X}_{\mathrm{even}} \oplus \mathcal{X}_{\mathrm{odd}}$, then $\mathcal{X}'_{\mathrm{Rad}} = \mathcal{X}'_{\mathrm{even}}$ and $\mathrm{P}^*_{\mathrm{Rad}} = \mathrm{P}_{\mathrm{Rad}} = \mathrm{P}_{\mathrm{even}}$, which is the self-adjoint projector that extracts the even part of a function.

4. Radon-Domain Sampling Functionals

The canonical sampling functional that acts on continuous functions expressed in hyper-spherical coordinates is the shifted hyper-spherical Dirac distribution $\delta_{\boldsymbol{z}_0} \in \mathcal{M}(\mathbb{R} \times \mathbb{S}^{d-1})$ with $\boldsymbol{z}_0 = (t_0, \boldsymbol{\xi}_0) \in \mathbb{R} \times \mathbb{S}^{d-1}$ and $\|\delta_{\boldsymbol{z}_0}\|_{\mathcal{M}} = 1$, which also happens to be a "formal" filtered projection of $\delta(\boldsymbol{\xi}_0^{\mathsf{T}}\boldsymbol{x} - t_0)$.

We now show how we can describe $[\delta_{\boldsymbol{z}_0}] \in \mathcal{S}'_{\rm Rad}$ by its unique representer $e_{\boldsymbol{z}_0} : g \mapsto g(\boldsymbol{z}_0)$ in $\mathcal{M}_{\rm Rad} = (C_{0,{\rm Rad}})'$. We start with a theoretical investigation where $e_{\boldsymbol{z}_0}$ is characterized indirectly through its functional properties. We then provide an explicit construction that allows us to identity $e_{\boldsymbol{z}_0}$ as the limit of a normalized Radon-compatible distribution whose unit mass gets concentrated at $\boldsymbol{z} = \pm\boldsymbol{z}_0$.

4.1 Abstract Characterization of Radon-Compatible Diracs

Since $\mathcal{S}_{\rm Rad} \subseteq \mathcal{S}_{\rm even}(\mathbb{R}\times\mathbb{S}^{d-1})$ (see Theorem 3), we have that $C_{0,{\rm Rad}} \subseteq C_{0,{\rm even}}(\mathbb{R}\times\mathbb{S}^{d-1})$, which implies that all functions $g \in C_{0,{\rm Rad}}$ are continuous and even. This ensures that the evaluation functional $e_{\boldsymbol{z}_0} : g \mapsto \langle e_{\boldsymbol{z}_0}, f\rangle = g(\boldsymbol{z}_0)$ is well-defined for any $\boldsymbol{z}_0 \in \mathbb{R}\times\mathbb{S}^{d-1}$. We now prove that $e_{\boldsymbol{z}_0}$ is a continuous linear functional on $C_{0,{\rm Rad}}$—that is, $e_{\boldsymbol{z}_0} \in (C_{0,{\rm Rad}})' = \mathcal{M}_{\rm Rad}$ (the space of Radon-compatible measures)—and establish its basic properties, which are compabible with those of the Dirac distribution. By the same token, we get a characterization of the extreme points of the unit ball in $\mathcal{M}_{\rm Rad}$.

Theorem 9 (Properties of $e_{\boldsymbol{z}_0}$) *The Radon-domain functionals $e_{\boldsymbol{z}_0} : C_{0,{\rm Rad}} \to \mathbb{R}$ with $\boldsymbol{z}_0 = (t_0, \boldsymbol{\xi}_0) \in \mathbb{R}\times\mathbb{S}^{d-1}$ have the following properties.*

1. *Definition (sampling at $\boldsymbol{z}_0$)*
$$\forall \phi \in C_{0,{\rm Rad}} : \quad \langle e_{(t_0,\boldsymbol{\xi}_0)}, \phi\rangle_{\rm Rad} = \phi(t_0, \boldsymbol{\xi}_0). \tag{40}$$
2. *Symmetry:* $e_{\boldsymbol{z}_0} = e_{-\boldsymbol{z}_0}$.
3. *Continuity:* $e_{\boldsymbol{z}_0} \in \mathcal{M}_{\rm Rad}$ *with* $\|e_{\boldsymbol{z}_0}\|_{\mathcal{M}_{\rm Rad}} = \sup\limits_{\phi\in C_{0,{\rm Rad}}:\ \|\phi\|_{L_\infty}\le 1} \langle e_{\boldsymbol{z}_0}, \phi\rangle = 1.$
4. *Let $\{\boldsymbol{z}_k\}$ be any finite set of distinct points. Then,* $\|\sum_k a_k e_{\boldsymbol{z}_k}\|_{\mathcal{M}_{\rm Rad}} = \sum_k |a_k|$.
5. $\mathrm{R}^*\{e_{(t_0,\boldsymbol{\xi}_0)}\}(\boldsymbol{x}) = \delta(\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x} - t_0) \ \Leftrightarrow\ e_{(t_0,\boldsymbol{\xi}_0)} = \mathrm{K}_{\rm rad}\mathrm{R}\{\delta(\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x} - t_0)\}$ *in* $\mathcal{M}_{\rm Rad}$.
6. *If* $e \in \mathrm{Ext}B_{\mathcal{M}_{\rm Rad}}$*, then* $e = \pm e_{\boldsymbol{z}_k}$ *for some* $\boldsymbol{z}_k \in \mathbb{R}\times\mathbb{S}^{d-1}$.

Proof Let ι be the canonical inclusion from $C_{0,{\rm Rad}}$ into $C_0(\mathbb{R}\times\mathbb{S}^{d-1})$ and $\iota^* : \mathcal{M}(\mathbb{R}\times\mathbb{S}^{d-1}) \to \mathcal{M}_{\rm Rad}$ the canonical projection from $\mathcal{M}(\mathbb{R}\times\mathbb{S}^{d-1})$ into $\mathcal{M}_{\rm Rad}$. For any $\boldsymbol{z}_0 \in \mathbb{R}\times\mathbb{S}^{d-1}$, we can then identify $e_{\boldsymbol{z}_0} = \iota^*\{\delta_{\boldsymbol{z}_0}\}$ where $\delta_{(t_0,\boldsymbol{\xi}_0)} \in \mathcal{M}(\mathbb{R}\times\mathbb{S}^{d-1})$ is the hyperspherical Dirac distribution. Indeed, for all $\phi \in C_{0,{\rm Rad}}$, we have that $\langle e_{\boldsymbol{z}_0}, \phi\rangle = \langle \delta_{\boldsymbol{z}_0}, \iota\{\phi\}\rangle = \langle \delta_{\boldsymbol{z}_0}, \phi\rangle = \phi(\boldsymbol{z}_0)$, from which we readily deduces Items 1, 2, and 3.

To prove Item 4, we consider $f = \sum_{k=1}^K a_k e_{\boldsymbol{z}_k}$, which is such that $\|f\|_{\mathcal{M}_{\rm Rad}} \le \sum_{k=1}^K |a_k| = \|\boldsymbol{a}\|_{\ell_1}$, by the triangle inequality. In addition, since the $\boldsymbol{z}_k$ are distinct, there exits some $\epsilon_0 > 0$ such that $\|\boldsymbol{z}_k - \boldsymbol{z}_{k'}\| < \epsilon_0$ for all $k \ne k'$. To prove that $\|f\|_{\mathcal{M}_{\rm Rad}} = \|\boldsymbol{a}\|_{\ell_1}$, we shall construct a conjugate function $f^* \in \mathcal{S}_{\rm Rad} \subset C_{0,{\rm Rad}}$ with $\|f^*\|_{L_\infty} = 1$. To that end, we use the functions
$$\phi_{\epsilon,\boldsymbol{x}_k} = \frac{1}{d_\epsilon(0,\boldsymbol{e}_1)} d_\epsilon(t - \boldsymbol{\xi}^\mathsf{T}\boldsymbol{x}_k, \mathbf{U}_k\boldsymbol{\xi}) \in \mathcal{S}_{\rm rad},$$
where $d_\epsilon(t,\boldsymbol{\xi})$ is specified by (44) and where $\boldsymbol{x}_k \in \mathbb{R}^d$ and the rotation matrix $\mathbf{U}_k \in \mathbb{R}^{d\times d}$ are chosen such that $\boldsymbol{\xi}_k^\mathsf{T}\boldsymbol{x}_k = t_k$ and $\boldsymbol{e}_1 = \mathbf{U}_k\boldsymbol{\xi}_k$. The function $\phi_{\epsilon,\boldsymbol{x}_k}$ with $\epsilon \in (0,1)$ is symmetric,

nonnegative and bounded by 1: It achieves its maximum at $\boldsymbol{z} = \boldsymbol{z}_k$ and is decreasing toward zero as $\boldsymbol{z}$ moves away from $\pm\boldsymbol{z}_k$, the speed of decay becoming arbitrarily fast as $\epsilon \to 0$ (see Figure 1). Consequently, one can always find some $\epsilon > 0$ such that $|\phi_{\epsilon,\boldsymbol{x}_k}(\boldsymbol{z})| < 1/K$ for all $\boldsymbol{z} \in \mathbb{R} \times \mathbb{S}^{d-1}$ with $\|\boldsymbol{z} \pm \boldsymbol{z}_k\| > \epsilon_0$. Then, $f^* = \sum_{k=1}^{K} \mathrm{sign}(a_k)\phi_{\epsilon,\boldsymbol{x}_k} \in \mathcal{S}_{\mathrm{Rad}}$ is such that $\|f^*\|_{L_\infty} = 1$ and $\langle f^*, f\rangle = \sum_{k=1} |a_k|$. The latter implies that $\|f\|_{\mathcal{M}_{\mathrm{Rad}}} \geq \|\boldsymbol{a}\|_{\ell_1}$, which proves the claim.

Item 5: Theorem 8 with $\mathcal{X} = C_0(\mathbb{R} \times \mathbb{S}^{d-1})$ ensures that the adjoint pair of operators $\mathrm{K}_{\mathrm{rad}}\mathrm{R}^* : C_{0,\mathrm{Rad}} \to \mathcal{Y}$ and $\mathrm{K}_{\mathrm{rad}}\mathrm{R} : \mathcal{Y}' \to \mathcal{M}_{\mathrm{Rad}}$ are continuous. This Banach setting also allows us to specify the corresponding back-projection operator $\mathrm{R}^* = (\mathrm{K}_{\mathrm{rad}}\mathrm{R})^{-1} : \mathcal{M}_{\mathrm{Rad}} \to \mathcal{Y}'$ by extending the scope of Definition (19) for $\varphi \in \mathcal{Y} = \mathrm{R}^*\mathrm{K}_{\mathrm{rad}}(C_{0,\mathrm{Rad}})$. We then use the same manipulations as in (23) with $\phi \in C_{0,\mathrm{Rad}}$ (resp., $\varphi \in \mathcal{Y}$) to prove that: (i) $\delta(\boldsymbol{\xi}_0^{\mathsf{T}}\boldsymbol{x} - t_0) = \mathrm{R}^*\{e_{(t_0,\boldsymbol{\xi}_0)}\} \in \mathcal{Y}'$, and (ii) $\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{\delta(\boldsymbol{\xi}_0^{\mathsf{T}}\boldsymbol{x} - t_0)\} = e_{(t_0,\boldsymbol{\xi}_0)}$ in $\mathcal{M}_{\mathrm{Rad}}$.

Item 6: The abstract interpretation of Items 1 and 2 is that the evaluation functionals on $C_{0,\mathrm{Rad}}$ are spanned (with a double covering) by $e_{\boldsymbol{z}}$ with $\boldsymbol{z} \in \mathcal{Z} \triangleq \mathbb{R} \times \mathbb{S}^{d-1}$. Since $C_{0,\mathrm{Rad}}$ is a closed subspace of $C_0(\mathcal{Z})$, we can invoke Lemma 17 in the Appendix, which tells us that the extreme points of the unit ball in $\mathcal{M}_{\mathrm{Rad}} = \big(C_{0,\mathrm{Rad}}\big)'$ are all of the form $\pm e_{\boldsymbol{z}}$ for some $\boldsymbol{z} \in \mathcal{Z}$.

■

4.2 Constructive Description of Radon-Compatible Dirac

In complement to the abstract characterization of $e_{\boldsymbol{z}_0}$ in Theorem 9, we now describe the underlying distribution concretely. We consider the d-dimensional Gaussian density function

$$g_\epsilon(\boldsymbol{x}) = \frac{\epsilon^{d-1}}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\big(\frac{x_1^2}{\epsilon^2} + \epsilon^2(x_2^2 + \cdots + x_d^2)\big)\right) \in \mathcal{S}(\mathbb{R}^d) \tag{41}$$

whose Fourier transform is

$$\hat{g}_\epsilon(\boldsymbol{\omega}) = \exp\left(-\frac{1}{2}\big(\epsilon^2\omega_1^2 + \frac{\omega_2^2 + \cdots + \omega_d^2}{\epsilon^2}\big)\right). \tag{42}$$

The parameter $\epsilon < 1$ controls the degree of ellipticity. When ϵ is small, $g_\epsilon(\boldsymbol{x})$ gets narrow along the x_1 axis, while it spreads out along the other directions. Setting $\boldsymbol{\omega} = \omega\boldsymbol{\xi}$, we rewrite (42) as

$$\hat{g}_\epsilon(\omega\boldsymbol{\xi}) = \exp\left(-\frac{\omega^2}{2}\sigma_\epsilon^2(\boldsymbol{\xi})\right) \text{ with } \sigma_\epsilon^2(\boldsymbol{\xi}) = \epsilon^2\xi_1^2 + \frac{\xi_2^2 + \cdots + \xi_d^2}{\epsilon^2}. \tag{43}$$

From (16), we obtain the Radon transform of g_ϵ as

$$d_\epsilon(t,\boldsymbol{\xi}) = \mathrm{R}\{g_\epsilon\}(t,\boldsymbol{\xi}) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2(\boldsymbol{\xi})}} \exp\left(-\frac{t^2}{2\sigma_\epsilon^2(\boldsymbol{\xi})}\right) \in \mathcal{S}_{\mathrm{Rad}}, \tag{44}$$

which is a radial Gaussian with a spherical dependence on the variance. For $0 < \epsilon < 1$, $d_\epsilon(t,\boldsymbol{\xi})$ attains its maximum at $(t,\boldsymbol{\xi}) = (0, \pm\boldsymbol{e}_1)$. As ϵ gets smaller, the maximum increases

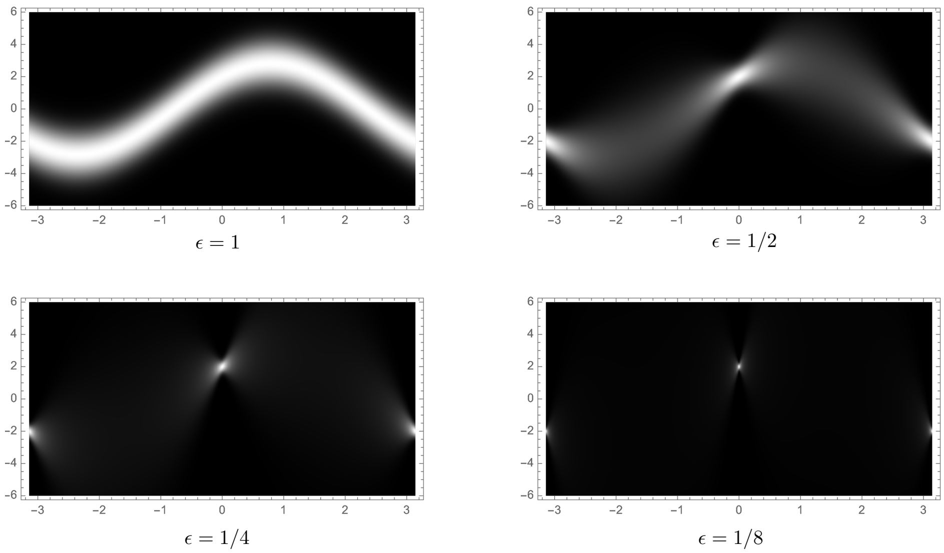

Figure 1: Localization effect of the parameter ϵ for the approximation of $e_{(2,\boldsymbol{e}_1)}$ for $d = 2$, displayed as the sinogram of the functions $d_\epsilon(t - \boldsymbol{\xi}^\mathsf{T}\boldsymbol{x}_0, \theta)$ with $\boldsymbol{\xi} = (\cos\theta, \sin\theta)$, $\boldsymbol{x}_0 = (2, 2)$, and $\epsilon = 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}$.

while the distribution becomes peakier and more and more localized around $\boldsymbol{z} = (0, \boldsymbol{e}_1)$. However, the integral of the function is preserved since $\int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} d_\epsilon(t, \boldsymbol{\xi}) \mathrm{d}t\mathrm{d}\boldsymbol{\xi} = \int_{\mathbb{S}^{d-1}} \mathrm{d}\boldsymbol{\xi} = \frac{2\pi^{d/2}}{\Gamma(d/2)}$ for any $\epsilon > 0$. This allows us to identify our Radon-compatible sampling functional as

$$e_{(t_0,\boldsymbol{\xi}_0)}(t, \boldsymbol{\xi}) = \lim_{\epsilon \to 0+} d_\epsilon(t - \boldsymbol{\xi}^\mathsf{T}\boldsymbol{x}_0, \mathbf{U}_0\boldsymbol{\xi}), \tag{45}$$

where $\mathbf{U}_0 \in \mathbb{R}^{d\times d}$ is a rotation matrix such that $\boldsymbol{e}_1 = \mathbf{U}_0\boldsymbol{\xi}_0$ and $\boldsymbol{x}_0 \in \mathbb{R}^d$ is such that $\boldsymbol{\xi}^\mathsf{T}\boldsymbol{x}_0 = t_0$. Examples of such functions for $d = 2$, $\boldsymbol{\xi}_0 = \boldsymbol{e}_1 = (1, 0)$, and $\boldsymbol{x}_0 = (2, 2)$ are shown in Figure 1.

While this construction reminds us of the description of a Dirac as the limit of a Gaussian distribution whose standard deviation tends to zero, there is one important difference: unlike a conventional Gaussian, the functions d_ϵ on the left hand side of (45) all satisfy the range conditions of the Radon transform, which are stated in Theorem 3 below.

5. Ridges Revisited

A 1D profile along the direction $\boldsymbol{e}_1 = (1, 0, \ldots, 0)$ is a generalized function of the form $r_{\boldsymbol{e}_1}(\boldsymbol{x}) = r(x_1) \times 1$ with $r \in \mathcal{S}'(\mathbb{R})$. Since the latter is separable, its generalized Fourier transform is

$$\mathcal{F}\{r_{\boldsymbol{e}_1}\}(\boldsymbol{\omega}) = \hat{r}(\omega_1) \prod_{k=2}^{d} 2\pi\delta(\omega_k) = \hat{r}(\omega_1)(2\pi)^{d-1}\delta(\omega_2, \ldots, \omega_d), \tag{46}$$

which is a weighted Dirac mass localized along the ω_1 axis. An equivalent formulation of (46) that involves test functions is

$$\forall\varphi\in\mathcal{S}(\mathbb{R}^d):\langle r_{\boldsymbol{e}_1},\varphi\rangle=\frac{1}{2\pi}\int_{\mathbb{R}}\hat{r}(\omega)\hat{\varphi}(\omega\boldsymbol{e}_1)\mathrm{d}\omega \tag{47}$$

where $\hat{\varphi}\in\mathcal{S}(\mathbb{R}^d)$ is the d-dimensional Fourier transform of φ. The argument remains valid when we rotate the coordinate system, which allows us to consider more general ridges of the form $r_{\boldsymbol{\xi}_0}(\boldsymbol{x})\stackrel{\triangle}{=}r(\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x})$; that is, 1D profiles along the direction $\boldsymbol{\xi}_0\in\mathbb{S}^{d-1}$.

5.1 Generalized Ridges

By identifying $\hat{\varphi}(\omega\boldsymbol{e}_1)$ in (47) as the 1D Fourier transform of $\mathrm{R}\{\varphi(\cdot,\boldsymbol{e}_1)\}$ and by substituting $\boldsymbol{e}_1$ by $\boldsymbol{\xi}_0$, we obtain the general signal-domain relation:

$$\forall\varphi\in\mathcal{S}(\mathbb{R}^d):\quad\langle r_{\boldsymbol{\xi}_0},\varphi\rangle=\langle r,\mathrm{R}\{\varphi\}(\cdot,\boldsymbol{\xi}_0)\rangle, \tag{48}$$

which will be referred to as the *ridge identity*. Under the assumption that r is a locally integrable function, we can establish (48) by making the change of coordinates $\boldsymbol{y}=\mathbf{U}\boldsymbol{x}$, where $\mathbf{U}\in\mathbb{R}^{d\times d}$ is any rotation matrix such that $y_1=\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x}$. We then rewrite the integral explicitly as

$$\begin{aligned}\int_{\mathbb{R}^d}\varphi(\boldsymbol{x})r_{\boldsymbol{\xi}_0}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}&=\int_{\mathbb{R}^d}\varphi(\boldsymbol{y})r(y_1)\mathrm{d}y_1\dots\mathrm{d}y_d\\&=\int_{\mathbb{R}}\underbrace{\left(\int_{\mathbb{R}^{d-1}}\varphi(\boldsymbol{y})\mathrm{d}y_2\dots\mathrm{d}y_d\right)}_{\mathrm{R}\varphi(y_1,\boldsymbol{\xi}_0)}r(y_1)\mathrm{d}y_1.\end{aligned}$$

Otherwise, when $r\in\mathcal{S}'(\mathbb{R})$ has no pointwise interpretation, we simply use (48) as definition for the ridge distribution $r_{\boldsymbol{\xi}_0}\in\mathcal{S}'(\mathbb{R}^d)$, which is legitimate since $\mathrm{R}\{\varphi\}(\cdot,\boldsymbol{\xi}_0)\in\mathcal{S}(\mathbb{R})$.

Note that the special case of (48) with $r(t)=\mathrm{e}^{-\mathrm{j}\omega t}$ and $r_{\boldsymbol{\xi}}(\boldsymbol{x})=\mathrm{e}^{-\mathrm{j}\omega\boldsymbol{\xi}^\mathsf{T}\boldsymbol{x}}=\mathrm{e}^{-\mathrm{j}\boldsymbol{\omega}^\mathsf{T}\boldsymbol{x}}$ yields the Fourier-slice theorem (16). Likewise, we can rely on the ridge identity (48) to delineate the range of the Radon transform. For instance, we obtain the second item in Theorem 3 by taking $r(t)=t^k$ and by defining

$$\Phi_k(\boldsymbol{\xi})=\int_{\mathbb{R}}\mathrm{R}\{\varphi\}(t,\boldsymbol{\xi})\;t^k\mathrm{d}t=\int_{\mathbb{R}^d}\varphi(\boldsymbol{x})(\boldsymbol{\xi}^\mathsf{T}\boldsymbol{x})^k\mathrm{d}\boldsymbol{x}=\sum_{|\boldsymbol{k}|=k}a_{\boldsymbol{k}}\boldsymbol{\xi}^{\boldsymbol{k}}$$

with $a_{\boldsymbol{k}}=k!\int_{\mathbb{R}^d}\frac{\boldsymbol{x}^{\boldsymbol{k}}}{\boldsymbol{k}!}\varphi(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$.

The most basic version of a ridge is $\delta(\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x}-t_0)$ with $r=\delta(\cdot-t_0)$, which is a Dirac ridge along $\boldsymbol{\xi}_0$ with offset t_0. Since the Fourier transform of such ridges is entirely localized along the ray $\{\boldsymbol{\omega}=\omega\boldsymbol{\xi}_0:\omega\in\mathbb{R}\}$, we can expect their Radon transform to vanish away from $\pm\boldsymbol{\xi}_0$. The latter can be readily identified as follows, where the square bracket notation $[g]$ with $g\in\mathcal{S}'(\mathbb{R}\times\mathbb{S}^{d-1})$ reminds us that the members of $\mathcal{S}'_{\mathrm{Rad}}$ (resp., of $\mathrm{K}_{\mathrm{rad}}\mathrm{R}\big(\mathcal{S}(\mathbb{R}^d)\big)'$) are equivalence classes of distributions.

Proposition 10 (Radon transform of ridge distributions) *Let* $(t_0, \boldsymbol{\xi}_0) = \boldsymbol{z}_0 \in \mathbb{R} \times \mathbb{S}^{d-1}$ *and* $r \in \mathcal{S}'(\mathbb{R})$. *Then,*

$$\begin{aligned}
\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{\delta(\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x} - t_0)\}(t, \boldsymbol{\xi}) &= [\delta(\cdot - t_0)\delta(\cdot - \boldsymbol{\xi}_0)] \quad \in \mathcal{S}'_{\mathrm{Rad}} \\
\mathrm{R}\{\delta(\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x})\}(t, \boldsymbol{\xi}) &= [q_d(t)\delta(\boldsymbol{\xi} - \boldsymbol{\xi}_0)] \quad \in \mathrm{K}_{\mathrm{rad}}\mathrm{R}\big(\mathcal{S}(\mathbb{R}^d)\big)' \\
\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{r(\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x})\}(t, \boldsymbol{\xi}) &= [r(t)\delta(\boldsymbol{\xi} - \boldsymbol{\xi}_0)] \quad \in \mathcal{S}'_{\mathrm{Rad}} \\
\mathrm{R}\{r(\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x})\}(t, \boldsymbol{\xi}) &= [(q_d * r)(t)\delta(\boldsymbol{\xi} - \boldsymbol{\xi}_0)] \quad \in \mathrm{K}_{\mathrm{rad}}\mathrm{R}\big(\mathcal{S}(\mathbb{R}^d)\big)'
\end{aligned}$$

where $q_d(t) = 2(2\pi)^{d-1}\mathcal{F}^{-1}\{1/|\omega|^{d-1}\}(t)$ *is the 1D impulse response of the Radon-domain inverse filtering operator* $\mathrm{K}_{\mathrm{rad}}^{-1}$.

Proof The fact that $\delta_{\boldsymbol{z}_0} = \delta(t - t_0)\delta(\boldsymbol{\xi} - \boldsymbol{\xi}_0)$ is a formal filtered projection of $\delta(\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x} - t_0)$ has already been mentioned in the text—it is a direct consequence of Definition (14).

To derive the third identity, we observe that, for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$, one has that

$$\begin{aligned}
\langle r(\cdot)\delta(\cdot - \boldsymbol{\xi}_0), \mathrm{R}\{\varphi\}\rangle_{\mathrm{Rad}} &= \int_{\mathbb{R}} r(t)\mathrm{R}\{\varphi\}(t, \boldsymbol{\xi}_0)\mathrm{d}t = \langle r, \mathrm{R}\{\varphi\}(\cdot, \boldsymbol{\xi}_0)\rangle \\
&= \int_{\mathbb{R}^d} r(\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x})\varphi(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \langle r(\boldsymbol{\xi}_0^\mathsf{T}\cdot), \mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\{\mathrm{R}\varphi\}\rangle,
\end{aligned}$$

where we have made use of (48). By setting $\phi = \mathrm{R}\{\varphi\} \in \mathcal{S}_{\mathrm{rad}}$, we then refer to (21) to deduce that $r(t)\delta(\boldsymbol{\xi} - \boldsymbol{\xi}_0)$ is a formal filtered projection of $r(\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x})$. This then also yields the second and fourth identities by substituting r with $q_d(\cdot - t_0)$ and $q_d * r$, respectively. The additional element there is $r(\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x}) = \mathrm{K}\mathrm{K}^{-1}\{r(\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x})\}(\boldsymbol{x}) = \mathrm{K}\{\tilde{r}(\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x})\}(\boldsymbol{x})$ with $\tilde{r}(t) = (q_d * r)(t)$, which is readily verified in the Fourier domain. ■

An equivalent form of the first identity in Proposition 10 is

$$\delta(\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x} - t_0) = \mathrm{R}^*\{\delta_{(t_0, \boldsymbol{\xi}_0)}\}(\boldsymbol{x}), \tag{49}$$

which results from $\mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\mathrm{R} = \mathrm{Id}$ on $\mathcal{S}'(\mathbb{R}^d)$. Likewise, when the back-projection in (49) is followed by an isotropic operator L whose radial frequency response is $\widehat{L}_{\mathrm{rad}} : \mathbb{R} \to \mathbb{R}$, we use the intertwining property to show that

$$\mathrm{L}\mathrm{R}^*\{\delta_{(t_0, \boldsymbol{\xi}_0)}\}(\boldsymbol{x}) = \mathrm{R}^*\{\mathrm{L}_{\mathrm{rad}}\delta_{(t_0, \boldsymbol{\xi}_0)}\}(\boldsymbol{x}) = r(\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x} - t_0) \tag{50}$$

where $r(t) = \mathcal{F}^{-1}\{\widehat{L}_{\mathrm{rad}}\}(t)$.

5.2 Connection with Prior Works

Most authors who use the Radon transform in connection with neural networks do not distinguish "formal" from "range-compatible" Radon transforms of distributions (Candès, 1999; Sonoda and Murata, 2017; Ongie et al., 2020; Parhi and Nowak, 2021). They bypass the difficulty by focusing their analysis on some appropriate subspace of $\mathcal{S}'_{\mathrm{Rad}}$ over which the backpropagation operator R^* is known to be invertible, for instance $\mathcal{M}_{\mathrm{even}}(\mathbb{R} \times \mathbb{S}^{d-1}) =$

$\mathcal{M}(\mathbb{P}^d)$ and/or $\mathcal{S}'_{\text{Liz,even}}$. To specify their norm for Radon-domain measures, Ongie et al. (2020) consider the subspace of test functions $\mathcal{S}_{\text{even}} = \{\phi \in \mathcal{S}(\mathbb{R}\times\mathbb{S}^d) : \phi(t,\boldsymbol{\xi}) = \phi(-t,-\boldsymbol{\xi})\}$ for which the validity of the inversion formula $\mathrm{K}_{\text{rad}}\mathrm{R}\mathrm{R}^* = \mathrm{Id}$ has been established by Solmon (1987). The caveat is that the functions $f \in \mathrm{R}^*(\mathcal{S}_{\text{even}})$ that are in the range of R^* can decay as badly as $O(1/\|\boldsymbol{x}\|)$ (Ramm and Katsevich, 2020, Corollary 3.1.1, p. 73), meaning that their "classical" Radon transform specified by (14) can be ill-defined. Parhi and Nowak (2021) follow a different path and identify $\mathcal{M}_{\text{even}}(\mathbb{R}\times\mathbb{S}^{d-1})$ as a subspace of the space of even Lizorkin distributions $\mathcal{S}'_{\text{Liz,even}}$, which is the topological dual of $\mathcal{S}_{\text{Liz,even}} = \{\phi \in \mathcal{S}_{\text{even}} : \int_{\mathbb{R}} \phi(t,\boldsymbol{\xi})t^k = 0, \forall k \in \mathbb{N}, \boldsymbol{\xi} \in \mathbb{S}^{d-1}\}$. Implicit in the calculation of Example 1 in (Ongie et al., 2020) is the property that

$$\mathrm{K}_{\text{rad}}\mathrm{R}\{\delta(\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x} - t_0)\}(t,\boldsymbol{\xi}) = \frac{1}{2}\Big(\delta(t-t_0)\delta(\boldsymbol{\xi}-\boldsymbol{\xi}_0) + \delta(t+t_0)\delta(\boldsymbol{\xi}+\boldsymbol{\xi}_0)\Big), \tag{51}$$

which needs to be related to the "abstract" description of $e_{\boldsymbol{z}_0}$ in Theorem 9. As it turns out, the two forms are equivalent. The abstract version, combined with the properties in Theorem 9, conveys the same information as (51). In complement is the concrete representation of $e_{(t_0,\boldsymbol{\xi}_0)}(t,\boldsymbol{\xi})$ given by (45), which gives us a sense of how and why the Radon-domain mass concentrates around the points $\pm\boldsymbol{z}_0$ as the Gaussian "blob" g_ϵ gets thinner along the primary axis and elongates in the perpendicular directions.

Even though the test functions used in the listed works are different from ours with $\mathcal{S}_{\text{Liz}} \subset \mathcal{S}_{\text{Rad}} \subset \mathcal{S}_{\text{even}}$, the approaches are reconciled by invoking the property that

$$C_{0,\text{even}} = \overline{(\mathcal{S}_{\text{even}}, \|\cdot\|_{L_\infty})} = \overline{(\mathcal{S}_{\text{Liz}}, \|\cdot\|_{L_\infty})} = \overline{(\mathcal{S}_{\text{Rad}}, \|\cdot\|_{L_\infty})} = C_{0,\text{Rad}} \tag{52}$$

where the domain of all underlying spaces is $(\mathbb{R}\times\mathbb{S}^{d-1})$ (Neumayer and Unser, 2022, Lemma 1). While Proposition 10 gives the filtered Radon transform of ridges with the greatest possible level of generality, the caveat is that $\mathrm{K}_{\text{rad}}\mathrm{R}\{r_{\boldsymbol{\xi}_0}\} \in \mathcal{S}'_{\text{Rad}}$ is an abstract equivalence class. As complement, we are providing the "concrete" version of the main result for the case where the profile is a measure.

Corollary 11 (Filtered Radon Transform of Ridge Measures) *Let $r_{\boldsymbol{\xi}_0} = r(\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x})$ be the ridge with profile $r \in \mathcal{S}'(\mathbb{R})$ and direction $\boldsymbol{\xi}_0 \in \mathbb{S}^{d-1}$. If $r \in \mathcal{M}(\mathbb{R})$, then the equality*

$$\mathrm{K}_{\text{rad}}\mathrm{R}\{r_{\boldsymbol{\xi}_0}\}(t,\boldsymbol{\xi}) = \frac{1}{2}\big(r(t)\delta(\boldsymbol{\xi}-\boldsymbol{\xi}_0) + r(-t)\delta(\boldsymbol{\xi}+\boldsymbol{\xi}_0)\big) \tag{53}$$

holds in $\mathcal{M}_{\text{Rad}} = \mathcal{M}_{\text{even}}(\mathbb{R}\times\mathbb{S}^{d-1})$.

Indeed, we know that $r(t)\delta(\boldsymbol{\xi}-\boldsymbol{\xi}_0)$ is a formal filtered Radon transform of $r_{\boldsymbol{\xi}_0}$ and that it is included in $\mathcal{M}(\mathbb{R}\times\mathbb{S}^{d-1})$ if $r \in \mathcal{M}(\mathbb{R})$. Moreover, Theorem 8 with $\mathcal{X} = C_0(\mathbb{R}\times\mathbb{S}^{d-1})$, together with (52), implies that $\mathcal{M}_{\text{Rad}} = (C_{0,\text{even}})' = \mathcal{M}_{\text{even}}$, whose Banach complement in $\mathcal{M}(\mathbb{R}\times\mathbb{S}^{d-1})$ is $\mathcal{M}_{\text{odd}}$. This ensures the validity of the second part of Theorem 8 with $\mathrm{P}^*_{\text{Rad}} = \mathrm{P}_{\text{even}}$, where $\mathrm{P}_{\text{even}}\{g\}(t,\boldsymbol{\xi}) = \frac{1}{2}\big(g(t,\boldsymbol{\xi}) + g(-t,-\boldsymbol{\xi})\big)$. Accordingly, we can identify $\mathrm{P}_{\text{even}}\{r(t)\delta(\boldsymbol{\xi}-\boldsymbol{\xi}_0)\}$ as the unique representer in $\mathcal{M}_{\text{Rad}}$ of $\mathrm{K}_{\text{rad}}\mathrm{R}\{r_{\boldsymbol{\xi}_0}\} = [r(t)\delta(\boldsymbol{\xi}-\boldsymbol{\xi}_0)] \in \mathcal{S}'_{\text{Rad}}$.

The argument also suggests that (53) is likely to be extendable to broader families of distributions. The condition for its validity is that $r(t)\delta(\boldsymbol{\xi}-\boldsymbol{\xi}_0)$ be included in a space $\mathcal{X}' = \big(\mathcal{X}_{\text{Rad}} \oplus \mathcal{X}^{\text{c}}_{\text{Rad}}\big)'$ with $\mathcal{X}_{\text{Rad}} = \overline{(\mathcal{S}_{\text{Rad}}, \|\cdot\|_{\mathcal{X}})} = \overline{(\mathcal{S}_{\text{even}}, \|\cdot\|_{\mathcal{X}})}$, so that the underlying projector can be readily identified as $\mathrm{P}^*_{\text{Rad}} = \mathrm{P}_{\text{even}} : \mathcal{X}' \to \mathcal{X}'_{\text{Rad}}$.

6. Variational Optimality of ReLU Networks

As application of the proposed formalism, we shall now link ReLU neural networks with functional optimization, revisiting the energy-minimization property uncovered in (Ongie et al., 2020) as well as the general variational-learning problem investigated in (Parhi and Nowak, 2021).

6.1 Learning with Radon-Domain Regularization

In order to state the relevant optimization problem, we consider the regularization operator

$$\Delta_{\mathrm{R}} = \mathrm{K}_{\mathrm{rad}}\mathrm{R}\Delta : \mathcal{M}_{\Delta_{\mathrm{R}}}(\mathbb{R}^d) \to \mathcal{M}(\mathbb{R} \times \mathbb{S}^{d-1}) \tag{54}$$

that was first proposed by Ongie et al. (2020, Lemma 2 p. 6), where Δ is the d-dimensional Laplace operator. The $\mathcal{M}$-norm (a.k.a. total variation) of this "Radonized" Laplacian is then used as regularization. Informally, the corresponding native space $\mathcal{M}_{\Delta_{\mathrm{R}}}(\mathbb{R}^d)$ is the largest subspace of continuous functions $f : \mathbb{R}^d \to \mathbb{R}$ such that: (i) f does not grow faster than a polynomial of degree 1, and (ii) the seminorm $\|\Delta_{\mathrm{R}} f\|_{\mathcal{M}}$ is well-defined and finite. The latter is a Banach space whose structure will be made explicit after the statement of the main theorem.

Theorem 12 (Optimality of Shallow ReLU Networks) *Let $E : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a strictly convex loss function, $(\boldsymbol{x}_m, y_m) \in \mathbb{R}^d \times \mathbb{R}$ with $m = 1, \dots, M$ a given set of distinct data points, and $\lambda > 0$ some fixed regularization parameter. Then, for $M > d + 1$, the solution set*

$$S = \left\{ \arg \min_{f \in \mathcal{M}_{\Delta_{\mathrm{R}}}(\mathbb{R}^d)} \left(\sum_{m=1}^{M} E(y_m, f(\boldsymbol{x}_m)) + \lambda \|\Delta_{\mathrm{R}} f\|_{\mathcal{M}} \right) \right\}, \tag{55}$$

of the functional optimization problem is nonempty and weak compact. It is the weak* closure of the convex hull of its extreme points, which all take the form*

$$f_{\mathrm{ridge}}(\boldsymbol{x}) = b_0 + \boldsymbol{b}^\top \boldsymbol{x} + \sum_{k=1}^{K_0} a_k \mathrm{ReLU}(\boldsymbol{\xi}_k^\mathsf{T} \boldsymbol{x} - \tau_k) \tag{56}$$

with $(b_0, \boldsymbol{b}) \in \mathbb{R} \times \mathbb{R}^d$, and a number $K_0 < M$ of adaptive ridges with weight, direction, and offset parameters $(a_k, \boldsymbol{\xi}_k, \tau_k) \in \mathbb{R} \times \mathbb{S}^{d-1} \times \mathbb{R}$. The corresponding regularization cost, which is common to all solutions, is $\|\Delta_{\mathrm{R}} f_{\mathrm{ridge}}\|_{\mathcal{M}} = \sum_{k=1}^{K_0} |a_k|$.

The key here is that the search space $\mathcal{M}_{\Delta_{\mathrm{R}}}(\mathbb{R}^d)$ is isometrically isomorphic to $\mathcal{M}_{\mathrm{Rad}} \times \mathcal{P}_1$, where $\mathcal{M}_{\mathrm{Rad}} = \big(C_{0,\mathrm{Rad}}\big)'$ (see Section 3 with $\|\cdot\|_{\mathcal{X}} = \|\cdot\|_{L_\infty}$ and $\mathcal{X}_{\mathrm{Rad}} = C_{0,\mathrm{Rad}}$) and $\mathcal{P}_1 = \{b_{\boldsymbol{k}} m_{\boldsymbol{k}}\}_{|\boldsymbol{k}| \le 1}$ is the space of affine functions on $\mathbb{R}^d$ (see (6) in Section 2.2 with $n_0 = 1$). An equivalent representation of the latter is $\mathcal{P}_1 = \{b_0 + \boldsymbol{b}^\mathsf{T} \boldsymbol{x}\}$ with $b_0 = b_{\boldsymbol{0}} \in \mathbb{R}$ and $\boldsymbol{b} = (b_{\boldsymbol{e}_i}) \in \mathbb{R}^d$, which matches the leading term in (56).

The crucial element for this construction is the pseudoinverse operator

$$\Delta_{\mathrm{R}}^{\dagger} \stackrel{\vartriangle}{=} (\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}_1}) \Delta^{-1} \mathrm{R}^* : \mathcal{M}_{\mathrm{Rad}} \to \mathcal{M}_{\Delta_{\mathrm{R}}}(\mathbb{R}^d), \tag{57}$$

where $\mathrm{Proj}_{\mathcal{P}_1} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{P}_1$ is the projector defined by (12) with $n_0 = 1$. The native space is then given by

$$\begin{aligned}\mathcal{M}_{\Delta_\mathrm{R}}(\mathbb{R}^d) &= \Delta_\mathrm{R}^\dagger\big(\mathcal{M}_\mathrm{Rad}\big) \oplus \mathcal{P}_1 \\ &= \{\Delta_\mathrm{R}^\dagger\{w\} + p_0 : (w, p_0) \in \mathcal{M}_\mathrm{Rad} \times \mathcal{P}_1\} \end{aligned} \tag{58}$$

equipped with the composite norm induced by $\mathcal{M}_\mathrm{Rad} \times \mathcal{P}_1$. Equivalently, given the generic form of f in (58), it is possible to retrieve the components (w, p_0) with the help of suitable linear maps. Specifically, letting $f = \Delta_\mathrm{R}^\dagger\{w\} + p_0$, one verifies that

$$\begin{aligned}\Delta_\mathrm{R}\{f\} &= \Delta_\mathrm{R}\{\Delta_\mathrm{R}^\dagger w\} + \Delta_\mathrm{R}\{p_0\} \\ &= \mathrm{K}_\mathrm{rad}\mathrm{R}\Delta(\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}_1})\Delta^{-1}\mathrm{R}^*\{w\} + 0 \\ &= \mathrm{K}_\mathrm{rad}\mathrm{R}\Delta\Delta^{-1}\mathrm{R}^*\{w\} - \mathrm{K}_\mathrm{rad}\mathrm{R}\,\underbrace{\Delta\{\mathrm{Proj}_{\mathcal{P}_1}\Delta^{-1}\mathrm{R}^* w\}}_{0} \\ &= \mathrm{K}_\mathrm{rad}\mathrm{R}\mathrm{R}^*\{w\} = w \in \mathcal{M}_\mathrm{Rad} \qquad \text{(by Theorem 8, Item 1)}\end{aligned}$$

and

$$\begin{aligned}\mathrm{Proj}_{\mathcal{P}_1}\{f\} &= \mathrm{Proj}_{\mathcal{P}_1}\{\Delta_\mathrm{R}^\dagger w\} + \mathrm{Proj}_{\mathcal{P}_1}\{p_0\} \\ &= \mathrm{Proj}_{\mathcal{P}_1}(\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}_1})\Delta^{-1}\mathrm{R}^*\{w\} + p_0 = 0 + p_0 \in \mathcal{P}_1,\end{aligned}$$

where the annihilation of the w-component follows from the idempotence of the projector.

Since we have not yet identified $\mathcal{M}_{\Delta_\mathrm{R}}(\mathbb{R}^d)$ as a "concrete" space of functions, we also need to interpret the sample values of f in (55) as linear functionals; that is, $f(\boldsymbol{x}_m) = \langle \delta(\cdot - \boldsymbol{x}_m), f\rangle$. This is enabled by the property that, for any $\boldsymbol{x}_m \in \mathbb{R}^d$, the sampling functional $\delta(\cdot - \boldsymbol{x}_m) : \mathcal{M}_{\Delta_\mathrm{R}}(\mathbb{R}^d) \to \mathbb{R}$ is weak*-continuous (see remark after Theorem 13).

Proof [Proof of Theorem 12] Theorem 8 with $\mathcal{X} = C_0(\mathbb{R} \times \mathbb{S}^{d-1})$ ensures that the back-projection operator R^* is invertible on $\mathcal{X}'_\mathrm{Rad} = \mathcal{M}_\mathrm{Rad} = (C_{0,\mathrm{Rad}})'$. This is the fundamental ingredient that makes $\mathcal{M}_{\Delta_\mathrm{R}}(\mathbb{R}^d)$ isometrically isomorphic to $\mathcal{M}_\mathrm{Rad} \times \mathcal{P}_1$ via the reversible mapping $w = \Delta_\mathrm{R}\{f\} \in \mathcal{M}_\mathrm{Rad}$ and $p_0 = \mathrm{Proj}_{\mathcal{P}_1}\{f\} \in \mathcal{P}_1$. This equivalent representation of f enables us to derive the result as a corollary of the third case of the abstract representer theorem for direct sums in (Unser and Aziznejad, 2022, Theorem 3). This representer theorem gives the generic form of the extreme points of the solution set S as $f_\mathrm{extreme} = p_0 + \sum_{k=1}^{K_0} a_k e_k$, where $p_0 \in \mathcal{P}_1$ is a null-space component and where the e_k are extreme points of the unit ball $B_{\mathcal{U}'}$ of the primary-component space $\mathcal{U}' = \Delta_\mathrm{R}^\dagger\big(\mathcal{M}_\mathrm{Rad}\big)$. Based on the form of the extreme points of $\mathcal{M}_\mathrm{Rad}$ given by Theorem 9 and the property that $\mathrm{Ext}B_{\mathcal{U}'} = \Delta_\mathrm{R}^\dagger\big(\mathrm{Ext}B_{\mathcal{M}_\mathrm{Rad}}\big)$ (since $\Delta_\mathrm{R}^\dagger$ is an isometry), we deduce that any $e_k \in \mathrm{Ext}B_{\mathcal{U}'}$ can be written as

$$\begin{aligned} e_k &= \pm\Delta_\mathrm{R}^\dagger\{e_{(t_k,\boldsymbol{\xi}_k)}\} = (\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}_1})\Delta^{-1}\mathrm{R}^*\{e_{(t_k,\boldsymbol{\xi}_k)}\} \\ &= \pm(\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}_1})\{\tfrac{1}{2}|\boldsymbol{\xi}_k^\mathsf{T} \cdot - t_k|\} \\ &= \pm\tfrac{1}{2}|\boldsymbol{\xi}_k^\mathsf{T} \cdot - t_k| \pm p_k \quad \text{with} \quad p_k \in \mathcal{P}_1,\end{aligned}$$

where we have used (50) with $\mathrm{L} = \Delta^{-1}$ and $\widehat{L}_{\mathrm{rad}}(\omega) = 1/\omega^2$ to evaluate the back-projection. Since $\frac{1}{2}|t| = (t)_+ - \frac{1}{2}t$, we can also write e_k as

$$e_k = \pm(\boldsymbol{\xi}_k^\top \cdot -t_k)_+ \pm \tilde{p}_k,$$

where $\tilde{p}_k = p_k - \frac{1}{2}\left(\boldsymbol{\xi}_k^\top \cdot +t_k\right) \in \mathcal{P}_1$. We then obtain (56) by forming the linear combination of extreme points and by grouping all polynomial components in a single term $b_0 + \boldsymbol{b}^\mathsf{T}\boldsymbol{x}$. Since $\Delta_\mathrm{R}\{f_{\mathrm{ridge}}\} = \sum_{k=1}^{K_0} a_k e_{(t_k,\boldsymbol{\xi}_k)}$, we also deduce that $\|\Delta_\mathrm{R}\{f_{\mathrm{ridge}}\}\|_{\mathcal{M}} = \sum_{k=1}^{K_0} |a_k|$ by invoking Theorem 9. ■

6.2 Concrete Characterization of the Native Space

As complement to Theorem 12, we now make the construction of $\mathcal{M}_{\Delta_\mathrm{R}}(\mathbb{R}^d)$ explicit by identifying the parent space $\mathcal{M}_{\mathrm{Rad}} \times \mathcal{P}_1$ and by providing the integral representation of the inverse operator $\Delta_\mathrm{R}^\dagger$ in (58). To that end, we first introduce the space $C_{0,\Delta_\mathrm{R}}(\mathbb{R}^d) = \Delta_\mathrm{R}^*\big(C_{0,\mathrm{even}}\big) \oplus \mathcal{P}_1'$ with $\mathcal{P}_1' = \mathrm{span}\{m_{\boldsymbol{k}}^*\}_{|\boldsymbol{k}|\le 1}$ and

$$C_{0,\mathrm{even}} = \{\phi : \phi \in C_0(\mathbb{R} \times \mathbb{S}^{d-1}) \text{ and } \phi(\boldsymbol{z}) = \phi(-\boldsymbol{z})\}.$$

The underlying operator $\Delta_\mathrm{R}^* = \Delta \mathrm{R}^* \mathrm{K}_{\mathrm{rad}}$ (the adjoint of Δ_R) is injective on $C_{0,\mathrm{even}}$ so that C_{0,Δ_R} can be equipped with the norm induced by the parent space $C_{0,\mathrm{even}} \times \mathcal{P}_1'$. Since $C_{0,\mathrm{even}} = C_{0,\mathrm{Rad}}$ (see (52)), $C_{0,\Delta_\mathrm{R}}(\mathbb{R}^d)$ is the predual of $\mathcal{M}_{\Delta_\mathrm{R}}(\mathbb{R}^d)$, which is itself isomorphic to $\mathcal{M}_{\mathrm{Rad}} \times \mathcal{P}_1 = \big(C_{0,\mathrm{Rad}} \times \mathcal{P}_1'\big)'$. The required left inverse of Δ_R^* is $\Delta_\mathrm{R}^{\dagger *} = \mathrm{R}\Delta^{-1}(\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}_1'}) : C_{0,\Delta_\mathrm{R}}(\mathbb{R}^d) \to C_{0,\mathrm{Rad}}$ (the adjoint of $\Delta_\mathrm{R}^\dagger$), where we have made the identification $(\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}_1})^* = (\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}_1'})$. These operators are such that $\Delta_\mathrm{R}^{\dagger *}\Delta_\mathrm{R}^* = \mathrm{Id}$ (left-inverse property) on $C_{0,\mathrm{even}}$ and $\Delta_\mathrm{R}^{\dagger *}\{p_0^*\} = 0$ for all $p_0^* \in \mathcal{P}_1'$. We then invoke the Riesz-Markov-Kakutani representation theorem (Rudin, 1973) to identify the space of Radon-compatible measures as $\mathcal{M}_{\mathrm{Rad}} = \big(C_{\mathrm{Rad}}\big)' = \big(C_{\mathrm{even}}\big)' = \mathcal{M}_{\mathrm{even}} = \mathrm{P}_{\mathrm{even}}\big(\mathcal{M}(\mathbb{R} \times \mathbb{S}^{d-1})\big)$.

The next step is to obtain the explicit expression of $\Delta_\mathrm{R}^\dagger\{w\} = (\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}_1})\Delta^{-1}\mathrm{R}^*\{w\}$ for $w \in \mathcal{M}_{\mathrm{even}}$ and, by extension, for $w \in \mathcal{M}(\mathbb{R}\times\mathbb{S}^{d-1})$ because the null space of $\Delta_\mathrm{R}^\dagger : \mathcal{M}(\mathbb{R}\times \mathbb{S}^{d-1}) \to \mathcal{M}_{\Delta_\mathrm{R}}(\mathbb{R}^d)$ $\big($and of $\mathrm{R}^* : \mathcal{M}(\mathbb{R} \times \mathbb{S}^{d-1}) \to \mathcal{S}'(\mathbb{R}^d)\big)$ is precisely the complementary space $\mathcal{M}_{\mathrm{Rad}}^{\mathrm{c}} = \mathcal{M}_{\mathrm{odd}}$ (see second part of Theorem 8). A direct calculation yields

$$\Delta_\mathrm{R}^\dagger\{w\}(\boldsymbol{x}) = \int_{\mathbb{R}} \int_{\mathbb{S}^{d-1}} h(\boldsymbol{x}; t, \boldsymbol{\xi}) w(t, \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi} \mathrm{d}t, \tag{59}$$

with a "Schwartz kernel" $h : \mathbb{R}^d \times (\mathbb{R} \times \mathbb{S}^{d-1}) \to \mathbb{R}$ that is given by

$$h(\boldsymbol{x}; t, \boldsymbol{\xi}) = \tfrac{1}{2}|\boldsymbol{\xi}^\mathsf{T}\boldsymbol{x} - t| - \sum_{|\boldsymbol{k}|\le 1} \frac{\boldsymbol{x}^{\boldsymbol{k}}}{\boldsymbol{k}!} q_{\boldsymbol{k}}(t, \boldsymbol{\xi}), \tag{60}$$

where $q_{\boldsymbol{k}}(t, \boldsymbol{\xi}) = \langle \frac{1}{2}|\boldsymbol{\xi}^\mathsf{T} \cdot -t|, m_{\boldsymbol{k}}^* \rangle$ and where the dual basis $(m_{\boldsymbol{k}}^*)_{|\boldsymbol{k}|\le 1}$ is specified by (7). This formula can be further simplified to obtain an analytical characterization of the Schwartz kernel of $\Delta_\mathrm{R}^\dagger$.

Theorem 13 (Explicit characterization of $\Delta_{\mathrm{R}}^{\dagger}$) *The generalized impulse response* $h(\boldsymbol{x};t,\boldsymbol{\xi}) = \Delta_{\mathrm{R}}^{\dagger}\{\delta_{(t,\boldsymbol{\xi})}\}(\boldsymbol{x})$ *of the inverse operator defined by* (57) *is given by*

$$h(\boldsymbol{x};t,\boldsymbol{\xi}) = \tfrac{1}{2}|t-\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}| - (\kappa_{\mathrm{rad}} * \tfrac{1}{2}|\cdot|)(t) + (\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x})(\kappa_{\mathrm{rad}} * \tfrac{1}{2}\mathrm{sign})(t). \tag{61}$$

Moreover, there exists a constant $C \geq 1$ *such that*

$$\big|h(\boldsymbol{x};t,\boldsymbol{\xi})\big| \leq C(1+\|\boldsymbol{x}\|), \tag{62}$$

while $h(\boldsymbol{x}_0;\cdot,\cdot) = \Delta_{\mathrm{R}}^{\dagger *}\{\delta(\cdot-\boldsymbol{x}_0)\} \in C_{0,\mathrm{even}}(\mathbb{R}\times\mathbb{S}^{d-1})$ *for any* $\boldsymbol{x}_0 \in \mathbb{R}^d$.

Proof One obtains (61) by invoking the ridge identity (48) together with (31) to evaluate $\langle \frac{1}{2}|\boldsymbol{\xi}^{\mathsf{T}}\cdot - t|, (-1)^{\boldsymbol{k}}\partial^{\boldsymbol{k}}\kappa_{\mathrm{iso}}\rangle$ and by then regrouping the first-order correction terms in (60).

To investigate the growth behavior of h, we first consider the non-mollified kernel

$$h_0(\boldsymbol{x};t,\boldsymbol{\xi}) = \tfrac{1}{2}|t-\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}| - \tfrac{1}{2}|t| + \tfrac{1}{2}(\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x})\mathrm{sign}(t), \tag{63}$$

which is obtained by setting $\kappa_{\mathrm{rad}} = \delta$ in (61). If we fix $t,\boldsymbol{\xi}$, then $\boldsymbol{x} \mapsto h_0(\boldsymbol{x};t,\boldsymbol{\xi})$ grows like $O(|\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}|)$ with the correction terms contributing a first-order polynomial. The effect of the correction is more radical along the radial variable t, as it neutralizes the growth of the leading term $\frac{1}{2}|t-\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}| = O(|t|)$ and produces a profile $t \mapsto h_0(\boldsymbol{x};t,\boldsymbol{\xi})$ that is compactly supported with a maximal value of $|\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}|$ at the origin. This remarkable behavior is illustrated in Figure 2. Since $|\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}| \leq \|\boldsymbol{x}\|$ (Cauchy-Schwarz), it also holds that $|h_0(\boldsymbol{x};t,\boldsymbol{\xi})| \leq \|\boldsymbol{x}\|$. To show that this growth profile applies to h as well, we rewrite the kernel as

$$h(\boldsymbol{x};t,\boldsymbol{\xi}) = g_0(t-\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}) + \big(\kappa_{\mathrm{rad}} * h_0(\boldsymbol{x};\cdot,\boldsymbol{\xi})\big)(t) \tag{64}$$

with $g_0(t) = (1-\kappa_{\mathrm{rad}}) * \frac{1}{2}|t|$. Next, we observe that the Fourier transform of g_0, $\widehat{g}_0(\omega) = \frac{1-\widehat{\kappa}_{\mathrm{rad}}(\omega)}{\omega^2}$, is bounded. Indeed, the numerator has a multiple-order zero at the origin that cancels the singularity in the denominator. We readily deduce that $\hat{g}_0 \in L_1(\mathbb{R})$, which implies that $g_0 \in C_0(\mathbb{R})$ with $|g_0(t)| \leq \|\hat{g}_0\|_{L_1}$ (Riemann-Lebesgue Lemma). As for the second convolution term, it is bounded by $\|\kappa_{\mathrm{rad}}\|_{L_1}\sup_{t\in\mathbb{R}}|h_0(\boldsymbol{x};t,\boldsymbol{\xi})| \leq \|\kappa_{\mathrm{rad}}\|_{L_1}\|\boldsymbol{x}\|$. This proves that (62) holds with $C = \|\hat{g}_0\|_{L_1} + \|\kappa_{\mathrm{rad}}\|_{L_1}$.

Since $\kappa_{\mathrm{rad}} \in \mathcal{S}(\mathbb{R})$ and $h_0(\boldsymbol{x};\cdot,\boldsymbol{\xi})$ is compactly supported, we have that $\kappa_{\mathrm{rad}} * h_0(\boldsymbol{x};\cdot,\boldsymbol{\xi}) \in \mathcal{S}(\mathbb{R})$ as well. The final claim then follows for the observation that the functions g_0, κ_{rad} and $h_0(\boldsymbol{x};\cdot,\cdot)$ that appear in (64) are all symmetric, bounded, and decaying towards zero. ■

An important consequence of the last statement in Theorem 13 is that $\delta(\cdot-\boldsymbol{x}_0) \in C_{0,\Delta_{\mathrm{R}}}(\mathbb{R}^d)$ for any $\boldsymbol{x}_0 \in \mathbb{R}^d$, which is equivalent to the sampling functional $\delta(\cdot-\boldsymbol{x}_0): f \mapsto f(\boldsymbol{x}_0)$ being weak*-continuous on $\mathcal{M}_{\Delta_{\mathrm{R}}} = \big(C_{0,\Delta_{\mathrm{R}}}\big)'$. The proof of the theorem is enlightening is that respect: The convolution with $\kappa_{\mathrm{rad}} \in \mathcal{S}(\mathbb{R})$ in (64) acts as a mollifier that wipes out the discontinuities of h_0. We also note that the whole argumentation (smoothing and pole cancellation) remains valid if κ_{rad} is substituted with any other symmetric kernel with a unit integral, such as a Gaussian with an arbitrarily small standard deviation.

The functional implications of Theorem 13 are as follows.

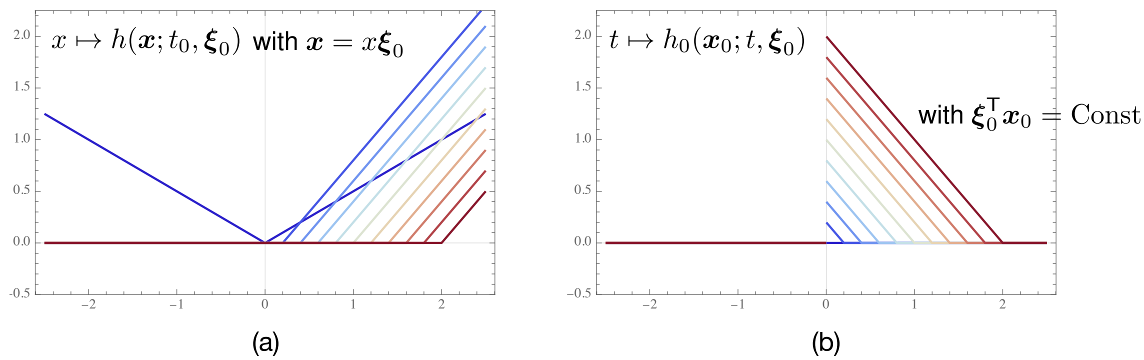

Figure 2: Visualization of $h_0(\boldsymbol{x}; t, \boldsymbol{\xi})$. (a) Plot of $\boldsymbol{x} \mapsto h_0(\boldsymbol{x}; t_0, \boldsymbol{\xi}_0)$ along the direction of the ridge with t_0 ranging from 0 (blue) to 2 (red). (b) Plot of $t \mapsto h_0(\boldsymbol{x}_0; t, \boldsymbol{\xi}_0)$ for $\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x}_0 = \mathrm{Const}$ varying between 0 and 2 (red).

Corollary 14 (Explicit characterization of the native space) *The statement $f \in \mathcal{M}_{\Delta_\mathrm{R}}(\mathbb{R}^d)$ is equivalent to*

$$f(\boldsymbol{x}) = b_0 + \mathbf{b}^\mathsf{T}\boldsymbol{x} + \int_\mathbb{R}\int_{\mathbb{S}^{d-1}} h(\boldsymbol{x}; t, \boldsymbol{\xi}) w(t, \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}\mathrm{d}t \tag{65}$$

for some $\boldsymbol{b} \in \mathbb{R}^d$, $b_0 \in \mathbb{R}$, and $w \in \mathcal{M}(\mathbb{R} \times \mathbb{S}^{d-1})$, with the integral representation being unique if $w = w_\mathrm{even} \in \mathcal{M}_\mathrm{even}$. In addition, the function specified by (65) is endowed with the following properties.

1. *Norm and seminorm*

$$\|f\|_{\mathcal{M}_{\Delta_\mathrm{R}}} = \|\Delta_\mathrm{R}\{f\}\|_\mathcal{M} + \|(\boldsymbol{b}, b_0)\|_2 < +\infty \tag{66}$$

$$\|\Delta_\mathrm{R}\{f\}\|_\mathcal{M} = \|\mathrm{P}_\mathrm{even}\{w\}\|_\mathcal{M} = \|w_\mathrm{even}\|_\mathcal{M} < +\infty. \tag{67}$$

2. *Linear growth*

$$\forall \boldsymbol{x} \in \mathbb{R}^d: \; |f(\boldsymbol{x})| \le C(1 + \|\boldsymbol{x}\|)\, \|f\|_{\mathcal{M}_{\Delta_\mathrm{R}}}. \tag{68}$$

3. *Continuity, with the function $f : \mathbb{R}^d \to \mathbb{R}$ also being differentiable with*

$$\forall \boldsymbol{x} \in \mathbb{R}^d: \; \|\boldsymbol{\nabla} f(\boldsymbol{x})\|_2 \le C'\|f\|_{\mathcal{M}_{\Delta_\mathrm{R}}}. \tag{69}$$

Proof The first statement and Item 1 are recapitulations, the latter re-expressing the isometric isomorphism with $\mathcal{M}_\mathrm{even} \times \mathcal{P}_1$. Item 2 is a direct consequence of (62), as

$$\begin{aligned}\left|\Delta_\mathrm{R}^\dagger\{f\}(\boldsymbol{x}) + \boldsymbol{b}^\mathsf{T}\boldsymbol{x} + b_0\right| &\le \int_\mathbb{R}\int_{\mathbb{S}^{d-1}} |h(\boldsymbol{x}; t, \boldsymbol{\xi})|\; |w(t, \boldsymbol{\xi})| \mathrm{d}\boldsymbol{\xi}\mathrm{d}t + \|\boldsymbol{b}\|_2\|\boldsymbol{x}\| + |b_0| \\ &\le C(1 + \|\boldsymbol{x}\|)\|w_\mathrm{even}\|_\mathcal{M} + \|(\boldsymbol{b}, b_0)\|_2.\end{aligned}$$

Likewise, we observe that the partial derivative

$$\partial_{x_i} h(\boldsymbol{x};t,\boldsymbol{\xi}) = \tfrac{-\xi_i}{2}\mathrm{sign}(t-\boldsymbol{\xi}^\mathsf{T}\boldsymbol{x}) - \tfrac{\xi_i}{2}(\kappa_{\mathrm{rad}} * \mathrm{sign})(t),$$

is bounded. We then apply the same technique to get the estimate $|\partial_{x_i} f(\boldsymbol{x})| \leq |b_i| + \|\kappa_{\mathrm{rad}}\|_{L_1}\|w_{\mathrm{even}}\|_{\mathcal{M}}$. ■

While the precise form of the kernel h depends on the specific choice of the dual basis, this has only an incidence of the definition of the norm, but not on the specification of $\mathcal{M}_{\Delta_{\mathrm{R}}}(\mathbb{R}^d)$ as a set. In other words, switching to another κ_{rad} will only change the way in which a given $f \in \mathcal{M}_{\Delta_{\mathrm{R}}}(\mathbb{R}^d)$ is decomposed into a sum of a primary term plus a polynomial.

6.3 Discussion

Our formulation of Theorem 12 owes a lot to the pioneering works of Ongie et al. (2020) and Parhi and Nowak (2021) (PN). Ours is merely a refinement of the results published by these authors together with a clarification of the underlying mathematics. The interesting outcome is that the solution of the functional-optimization problem in (55) can be implemented by a 2-layer ReLU network.

In their work which pulls together ideas from Unser et al. (2017) and Ongie et al. (2020), PN restrict the domain of the test functions to the so-called Lizorkin functions

$$\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d) = \{\varphi \in \mathcal{S}(\mathbb{R}^d) : \int_{\mathbb{R}^d} \boldsymbol{x}^{\boldsymbol{m}}\varphi(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = 0, \forall \boldsymbol{m} \in \mathbb{N}^d\}, \tag{70}$$

which are orthogonal to the polynomials. This choice is motivated by the property that the Radon transform is a homeomorphism $\mathrm{R} : \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d) \to \mathcal{S}_{\mathrm{Liz,even}}$, where $\mathcal{S}_{\mathrm{Liz,even}}$ denotes the even part of the Radon-domain Lizorkin space $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R} \times \mathbb{S}^{d-1})$ with $\mathrm{R}^*\mathrm{R}\mathrm{K}_{\mathrm{rad}} = \mathrm{Id}$ on $\mathcal{S}_{\mathrm{Liz,even}}$, as well as on $\mathcal{S}'_{\mathrm{Liz,even}}$, by duality (Helgason, 2011; Kostadinova et al., 2014).

While the adoption of this formalism leads to a well-defined functional-optimization problem, PN's derivation/interpretation of Lemmas 17, 18, and 21 is flawed because they implicitly assume that there is a systematic, one-to-one association between a "concrete" spline ridge $\rho_m(\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x} - t_0)$ with $\rho_m(t) = \frac{1}{2}\mathrm{sign}(t)\frac{t^{m-1}}{(m-1)!}$ and some abstract Lizorkin distribution $\rho_m(\boldsymbol{\xi}_0^\mathsf{T}\boldsymbol{x} - t_0) + \mathcal{P} \in \mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$, which is unlikely to be the case for the reasons exposed in the introduction. We have recent evidence that such an association can be made, but that it requires a specific polynomial correction that depends on the shift t_0 (Neumayer and Unser, 2022). That said, it remains that the main results and conclusions reported by Parhi and Nowak (2021) are qualitatively correct and in agreement with Theorem 12 (up to the mentioned technicalities). Also, the mathematical arguments proposed by these authors can easily be corrected/upgraded by extending their space of test functions to $\mathcal{S}_{\mathrm{Rad}}$ and by using our results in Theorems 8 and 9. The same holds true for PN's higher-order generalizations (ridge splines). In fact, PN's definition of the native space of the mth-order ridge splines is equivalent to that of $\mathcal{M}_{\Delta_{\mathrm{R}}}(\mathbb{R}^d)$ for $m = 2$.

Now, the one aspect where Theorem 12 improves upon Parhi and Nowak (2021, Theorem 1 with $m = 2$) is that it contains the characterization of the full solution set. The theorem tells us that *all extreme points* of the optimization problem in (55) have the same parametric form (56), which is a much stronger statement than the existence of one such neural-network-like solution. Ideally, one would like to identify the sparsest solution within the solution set, in other words, the one with the fewest neurons. While there is a direct algorithm that will find the sparsest solution for $d = 1$ (Debarre et al., 2022), it is not known yet if this can be generalized to a larger number of dimensions.

In addition to commuting with rotation, the regularization operator Δ_{R} defined by (54) inherits the scale invariance of the Laplacian. Specifically, we can use a direct Fourier calculation to show that

$$\forall \boldsymbol{x} \in \mathbb{R}^d : \quad \Delta\{f\}(s\boldsymbol{x}) = s^2 \Delta\{f(s\cdot)\}(\boldsymbol{x}) \tag{71}$$

$$\forall (t, \boldsymbol{\xi}) \in \mathbb{R} \times \mathbb{S}^{d-1} : \quad \Delta_{\mathrm{R}}\{f\}(st, \boldsymbol{\xi}) = s^2 \Delta_{\mathrm{R}}\{f(s\cdot)\}(t, \boldsymbol{\xi}) \tag{72}$$

for any scaling factor $s \in \mathbb{R}^+$. This, in turn, translates into the functional-learning problem in (55) being invariant to similarity transformations of the data points $\boldsymbol{x}_m$.

Proposition 15 *The regularization functional in* (55) *is translation-, scale-, and rotation-invariant in the sense that*

$$\|\Delta_{\mathrm{R}}\{f(s\mathbf{U} \cdot - \mathbf{b})\}\|_{\mathcal{M}} = s\|\Delta_{\mathrm{R}}\{f\}\|_{\mathcal{M}} \tag{73}$$

for any $f \in \mathcal{M}_{\Delta_{\mathrm{R}}}(\mathbb{R}^d)$, and any scaling factor $s \in \mathbb{R}^+$, offset $\mathbf{b} \in \mathbb{R}^d$, and rotation matrix $\mathbf{U} \in \mathbb{R}^{d \times d}$ with $\mathbf{U}^{\mathsf{T}}\mathbf{U} = \mathbf{I}$.

This is to say that any such transformation of the data characterized by a scale s can be accounted for via a proper rescaling of the regularization parameter $\lambda \to s\lambda$.

The form of the kernel given by (59) is compatible with (Parhi and Nowak, 2021, Lemma 21) and has been interpreted as an infinite-width neural network. There has been concern about the well-posedness of the generative model used in (Parhi and Nowak, 2021) and summarized by $f = \Delta_{\mathrm{R}}^{\dagger}\{w\}$ with $w \in \mathcal{M}_{\mathrm{Rad}} = \mathcal{M}_{\mathrm{even}}$ in the present formulation. This point is addressed explicitly by the stability bound (62) in Theorem 13, which is new to the best of our knowledge.

The availability of the integral formulation (65) also allows us to make the connection with the work of (Bartolucci et al., 2023). These authors consider the same type of generative model with some "tempered" kernel of the form $h_\beta(\boldsymbol{x}; t, \boldsymbol{\xi}) = \frac{1}{2}|\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x} - t|\,\beta(t)$, where $\beta(t) > 0$ is a weighting function (e.g., $\beta(t) = \frac{1}{1+|t|^{2+\epsilon}}$) that compensates the linear growth of the first factor[2](Bartolucci et al., 2023). In effect, this mechanism, whose stability is intrinsically guaranteed, reduces the size of the native space. Interestingly, the corresponding optimization problem admits the same form of solution—a neural network with one hidden ReLU layer (Bartolucci et al., 2023)—with the caveat that the underlying regularization is no longer translation-invariant. In this modified scenario, the optimal cost is $\|\Delta_{\mathrm{R}} f_{\mathrm{ridge}}\|_{\mathcal{M}_{1/\beta}} = \sum_{n=1}^{K_0} |a_k| \frac{1}{|\beta(\tau_k)|}$, which tends to favor smaller biases τ_k. This also means that one then loses the invariance to similarity transformations of the data (Proposition 15).

2. The first factor can also be replaced by $(\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x} - t)_+$ modulo some adjustments in β, as shown by the authors.

Appendix A. Direct-Sum Topologies

There are two standard ways to define direct sums: explicitly, via the use of projectors; or abstractly, via the use of quotient spaces. The two methods are equivalent whenever one can explicitly identify the underlying quotient space as a (concrete) complemented subspace of the original space.

A.1 Projectors

Let $\mathcal{X}$ be a topological vector space. A continuous linear operator $\mathrm{P} : \mathcal{X} \to \mathcal{X}$ with the property that $\mathrm{P} = \mathrm{P} \circ \mathrm{P} = \mathrm{P}^2$ (idempotence) on $\mathcal{X}$ is called a projection operator (Dunford and Schwartz, 1988, p 480). In particular, when $\mathcal{X}$ is a Banach space or a Fréchet space, the range $\mathcal{U} = \mathrm{P}(\mathcal{X})$ of P is necessarily a closed subspace of $\mathcal{X}$. In that case, $\mathrm{P} = \mathrm{Proj}_{\mathcal{U}}$ is a projector from $\mathcal{X}$ onto $\mathcal{U}$, and $\mathcal{X} = \mathcal{U} \oplus \mathcal{V}$ where $\mathcal{V} = \mathrm{ker}\mathrm{P}$ is the null space of P or, equivalently, the range of the complementary projector $\mathrm{Proj}_{\mathcal{V}} = (\mathrm{Id} - \mathrm{P})$.

More generally, when $\mathcal{X}$ is a topological vector space, the space $\mathrm{P}(\mathcal{X})$ equipped with the topology induced by $\mathcal{X}$ is a topological space as well, with the same properties as the original space $\mathcal{X}$ (e.g., completeness). Likewise, if $(\mathcal{X}, \mathcal{X}')$ is a dual pair of topological spaces, then so is $\big(\mathrm{P}(\mathcal{X}), \mathrm{P}^*(\mathcal{X}')\big)$, where $\mathrm{P}^* : \mathcal{X}' \to \mathcal{X}'$ is the dual projection operator.

A.2 Direct Sums

1) Direct-sum decomposition of a vector space $\mathcal{X}$: Let $\mathcal{U}$ and $\mathcal{V}$ be two (complementary) closed subspaces of $\mathcal{X}$. The notation $\mathcal{X} = \mathcal{U} \oplus \mathcal{V}$ indicates that every element $x \in \mathcal{X}$ has a unique decomposition as $x = u + v$ with $(u, v) \in \mathcal{U} \times \mathcal{V}$. The underlying projection operators are

$$x = u + v \mapsto \mathrm{Proj}_{\mathcal{U}}\{x\} = u$$
$$x \mapsto \mathrm{Proj}_{\mathcal{V}}\{x\} = (\mathrm{Id} - \mathrm{Proj}_{\mathcal{U}})\{x\} = v.$$

To summarize, given a (closed) subspace $\mathcal{U}$ of a normed vector space $\mathcal{X}$, the search of a complement $\mathcal{V}$ for $\mathcal{U}$ in $\mathcal{X}$ is equivalent to the search of a (continuous) projection operator P on $\mathcal{X}$ (with $\mathrm{P}^2 = \mathrm{P}$) whose range is $\mathcal{U}$. Then, $\mathcal{V} = \mathrm{Proj}_{\mathcal{V}}(\mathcal{X})$ with $\mathrm{Proj}_{\mathcal{V}} = (\mathrm{Id} - \mathrm{P})$.

2) Annihilator: Let $\mathcal{U}$ be a closed subset of $\mathcal{X}$. One then defines $\mathcal{U}^{\perp}$ as the annihilator of $\mathcal{U}$ in $\mathcal{X}'$, which is the subset

$$\mathcal{U}^{\perp} = \{f \in \mathcal{X}' : \langle f, u \rangle = 0 \text{ for all } u \in \mathcal{U}\} \subseteq \mathcal{X}'.$$

3) Dual space: The dual of the direct sum $\mathcal{X} = \mathcal{U} \oplus \mathcal{V}$ is $\mathcal{X}' = \mathcal{U}' \oplus \mathcal{V}'$, where $\mathcal{U}' = \mathrm{P}^*(\mathcal{X}') = \mathcal{V}^{\perp}$ and $\mathcal{V}' = \mathcal{U}^{\perp}$.

4) Quotient space: Under the assumption that $\mathcal{V}$ is a closed subset of $\mathcal{X}$, one defines the quotient space $\mathcal{X}/\mathcal{V}$ whose elements are equivalence classes (or cosets) denoted by $[x] = x + \mathcal{V}$. The corresponding quotient map $q : \mathcal{X} \to \mathcal{X}/\mathcal{V}$ is linear and its kernel (null space) is $\mathcal{V}$. When $\mathcal{X}$ is a Banach space, the quotient norm is

$$\|[x]\|_{\mathcal{X}/\mathcal{V}} = \inf_{v \in \mathcal{V}} \|x + v\|_{\mathcal{X}},$$

which measures the distance from x to $\mathcal{V}$. The quotient space $\mathcal{X}/\mathcal{V}$ equipped with the quotient norm is a Banach space as well. Moreover, there is a natural isomorphism between $(\mathcal{X}/\mathcal{V})'$ (the dual of the quotient of $\mathcal{X}$ by $\mathcal{V}$) and $\mathcal{V}^\perp$ (the annihilator of $\mathcal{V}$ in $\mathcal{X}'$), so that $(\mathcal{X}/\mathcal{V})' \hookrightarrow \mathcal{X}'$.

Also of relevance is the property that the bounded operators on $\mathcal{X}$ that annihilate the elements of $\mathcal{V}$ "factor through" $\mathcal{X}/\mathcal{V}$. Let $\mathrm{T} : \mathcal{X} \to \mathcal{Y}$ with $\mathcal{V} \subseteq \ker\mathrm{T}$. Then, there exists a unique linear operator $\mathrm{T}_q : \mathcal{X}/\mathcal{V} \to \mathcal{Y}$ such that $\mathrm{T}_q q(x) = \mathrm{T}(x)$ and $\|\mathrm{T}_q\| = \|\mathrm{T}\|$.

The kernel of any bounded operator $\mathrm{T} : \mathcal{X} \to \mathcal{Y}$ is a closed subspace of $\mathcal{X}$ (Markin, 2020, Proposition 4.2, p. 172). Hence, the quotient space $\mathcal{X}/\ker(\mathrm{T})$ is a vector space that is isomorphic to $\mathrm{T}(\mathcal{X})$.

Appendix B. Extreme Points

Definition 16 (Extreme Points) *Let C be a convex set of a Banach space $\mathcal{X}$. The extreme points of C are the points $x \in C$ such that, if there exist $x_1, x_2 \in C$ and $\theta \in (0, 1)$ such that $x = \theta x_1 + (1-\theta)x_2$, then it necessarily holds that $x_1 = x = x_2$. The set of these extreme points is denoted by* $\mathrm{Ext}(C)$.

We now present a classical result that gives the explicit form of the extreme points of the dual $\mathcal{X}'$ of any closed subspace $\mathcal{X} \subseteq C(\mathcal{Z})$, where $C(\mathcal{Z})$ is the space of continuous functions $z \mapsto f(z)$ on some compact Hausdorff space $\mathcal{Z}$ equipped with the norm $\|f\| = \sup_{z\in\mathcal{Z}} |f(z)|$.

Lemma 17 ((Dunford and Schwartz, 1988, p. 441)) *Let $\mathcal{X}$ be a closed linear manifold of the Banach space $C(\mathcal{Z})$ of all real continuous functions on the compact Hausdorff space $\mathcal{Z}$. For each $z \in \mathcal{Z}$, let the evaluation functional $e_z \in \mathcal{X}'$ be defined by*

$$\langle e_z, f\rangle = f(z), \quad f \in \mathcal{X}. \tag{74}$$

Then, every extreme point of the closed unit ball in $\mathcal{X}'$,

$$B_{\mathcal{X}'} = \{x^* \in \mathcal{X}' : \|x^*\|_{\mathcal{X}'} = \sup_{f\in\mathcal{X}:|f(z)|\leq 1} \langle x^*, f\rangle \leq 1\}, \tag{75}$$

is of the form $\pm e_z$ with $z \in \mathcal{Z}$. If $\mathcal{X} = C(\mathcal{Z})$, then the converse is true as well; that is, $\mathrm{Ext}B_{\mathcal{M}(\mathcal{Z})} = \{\pm e_z : z \in \mathcal{Z}\}$ *with* $\mathcal{M}(\mathcal{Z}) = \big(C(\mathcal{Z})\big)'$.

Lemma 17 generalizes to $C_0(\mathcal{Z})$, where $\mathcal{Z}$ is a locally compact Hausdorff space, which covers the case that is of interest to us: $\mathcal{Z} = \mathbb{R} \times \mathbb{S}^{d-1}$. The scenario $\mathcal{X} = C_0(\mathcal{Z})$ is well-known and covered, for instance, by (Bredies and Carioni, 2020, Proposition 4.1). The result can then be transferred to any closed subspace $\mathcal{X}$ by identifying $\mathcal{X}'$ as $\mathcal{M}(\mathcal{Z})/\mathcal{X}^\perp$ and applying the canonical projection operator with the help of (Bredies and Carioni, 2020, Lemma 3.2). Since the direct proof of the extended version of Lemma 17 is reasonably short, we are including it here to be self-contained.

Proof Let E be the set of all points in $\mathcal{X}'$ of the form $\pm e_z$ with $z \in \mathcal{Z}$. The space $\mathcal{X}'$ is equipped with its weak$*$ (or $\mathcal{X}$) topology for the Krein-Milman theorem to apply. As $\|e_z\|_{\mathcal{X}'} \leq 1$, $E \subseteq B_{\mathcal{X}'}$. Since $B_{\mathcal{X}'}$ is convex, weak*-compact and, hence, weak*-closed, the inclusion also holds for the closed convex hull, with $\mathrm{cch}E \subseteq B_{\mathcal{X}'}$. Next, we invoke a variant of the Hahn-Banach theorem (Rudin, 1991, Theorem 3.5, p. 59). For any $x^* \notin \mathrm{cch}E$, there exists a linear functional $f \in \mathcal{X}$ that separates $x^* \in \mathcal{X}'$ from the closed convex set $\mathrm{cch}E$. This means that there are a constant $c > 0$ and some $\epsilon > 0$ such that

$$\pm f(z) \leq c - \epsilon < c \leq \langle x^*, f \rangle$$

for all $z \in \mathcal{Z}$. Hence, $\|f\| \leq (c - \epsilon)$ which, when combined with $\|x^*\|_{\mathcal{X}'}\|f\| \geq c$, gives that $\|x^*\|_{\mathcal{X}'} > 1$. Thus, $\mathrm{cch}E \supseteq B_{\mathcal{X}'}$, from which we conclude that $\mathrm{cch}E = B_{\mathcal{X}'}$. Finally, since E is compact, the extreme points of $\mathrm{cch}E$ necessarily lie in E (see (Rudin, 1991, Milman's theorem, p. 76)).

For the converse implication, we invoke the Riesz-representation theorem, which allows us to represent any unit-norm functional on $C_0(\mathcal{Z})$ by a real-valued measure $\mu \in \mathcal{M}(\mathcal{Z})$ of total variation 1. If the support of μ consists of one point, it is a signed multiple of a Dirac mass. Otherwise, $\mathrm{supp}\mu$ contains two distinct points $z_1 \neq z_2$. Let $U, V \subset \mathcal{Z}$ be disjoint neighborhoods of z_1 and z_2. By the definition of the support, $|\mu|(U) > 0$ and $|\mu|(V) > 0$. Define $t = |\mu|(U)$, which is such that $0 < t < 1$. Now let $\lambda = t^{-1}\mu|_U$ and $\nu = (1-t)^{-1}\mu|_{U^c}$. Then, both λ and ν are unit-norm functionals and $\mu = t\lambda + (1-t)\nu$, which proves that μ is not extreme. ■

Acknowlegdments

The research was partially supported by the Swiss National Science Foundation under Grant 200020-184646. The author is thankful to Sebastian Neumayer, Shayan Aziznejad and Rahul Parhi for very helpful discussions. He would also like to thank the two anonymous reviewers for their insightful comments and concrete suggestions to improve the clarity of the exposition.

References

Mauricio A Alvarez, Lorenzo Rosasco, and Neil D Lawrence. Kernels for vector-valued functions: A review. *Foundations and Trends in Machine Learning*, 4(3):195–266, 2012.

Francis Bach. Breaking the curse of dimensionality with convex neural networks. *Journal of Machine Learning Research*, 18:1–53, 2017.

Andrew R Barron. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory*, 39(3):930–945, May 1993.

Francesca Bartolucci, Ernesto De Vito, Lorenzo Rosasco, and Stefano Vigogna. Understanding neural networks with reproducing kernel Banach spaces. *Applied and Computational Harmonic Analysis*, 62:194–236, 2023.

Christopher M Bishop. *Pattern Recognition and Machine Learning.* Springer, 2006.

Jan Boman and Filip Lindskog. Support theorems for the Radon transform and Cramér-Wold theorems. *Journal of Theoretical Probability*, 22(3):683–710, March 2009.

Kristian Bredies and Marcello Carioni. Sparsity of solutions for variational inverse problems with finite-dimensional data. *Calculus of Variations and Partial Differential Equations*, 59(14):26, 2020.

Emmanuel J Candès. Harmonic analysis of neural networks. *Applied and Computational Harmonic Analysis*, 6(2):197–218, 1999.

Emmanuel J Candès and David L Donoho. Ridgelets: A key to higher-dimensional intermittency? *Phil. Trans. R. Soc. Lond. A.*, pages 2495–2509, 1999.

Emmanuel J Candès and Justin Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969–985, 2007.

George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314, 1989.

Thomas Debarre, Quentin Denoyelle, Michael Unser, and Julien Fageot. Sparsest piecewise-linear representation of data. *Journal of Computational and Applied Mathematics*, 406: 1–30, 2022. paper no. 114044.

David L Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4): 1289–1306, 2006.

Nelson Dunford and Jacob T Schwartz. *Linear Operators, Part 1: General Theory*. John Wiley and Sons, 1988.

Michael Elad. *Sparse and Redundant Representations. From Theory to Applications in Signal and Image Processing*. Springer, 2010.

Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Springer, 2013.

Israel M Gelfand and Georgy Shilov. *Generalized Functions. Vol. 5. Integral Geometry and Representation Theory*. Academic Press, New York, 1966.

Sigurdur Helgason. *Integral Geometry and Radon Transforms*. Springer, 2011.

Alexander Hertle. Continuity of the Radon transform and its inverse on Euclidean space. *Mathematische Zeitschrift*, 184(2):165–192, 1983.

Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.

Sanja Kostadinova, Stevan Pilipović, Katerina Saneva, and Jasson Vindas. The ridgelet transform of distributions. *Integral Transforms and Special Functions*, 25(5):344–358, 2014.

Benjamin F Logan and Larry A Shepp. Optimal reconstruction of a function from its projections. *Duke Mathematical Journal*, 42(4):645–659, 1975.

Donald Ludwig. The Radon transform on Euclidean space. *Communications on Pure and Applied Mathematics*, 19(1):49–81, September 1966.

Wolodymyr R Madych. Summability and approximate reconstruction from Radon transform data. *Contemporary Mathematics*, 113:189–219, 1990.

Marat V Markin. *Elementary Operator Theory.* De Gruyter, 2020.

Noboru Murata. An integral representation of functions using three-layered networks and their approximation bounds. *Neural Networks*, 9(6):947–956, 1996.

Frank Natterer. *The Mathematics of Computed Tomography.* John Willey & Sons Ltd., 1984.

Sebastian Neumayer and Michael Unser. Explicit representations for Banach subspaces of Lizorkin distributions. *Preprint arXiv:2203.05312*, 2022.

Greg Ongie, Rebecca Willett, Daniel Soudry, and Nathan Srebro. A function space view of bounded norm infinite width ReLU nets: The multivariate case. *International Conference on Representation Learning (ICLR)*, 2020.

Rahul Parhi and Robert D Nowak. Banach space representer theorems for neural networks and ridge splines. *Journal of Machine Learning Research*, 22(43):1–40, 2021.

Allan Pinkus. *Ridge Functions.* Cambridge Tracts in Mathematics. Cambridge University Press, 2015.

Tomaso Poggio and Federico Girosi. Regularization algorithms for learning that are equivalent to multilayer networks. *Science*, 247(4945):978–982, 1990.

Alexander G Ramm. Inversion formula and singularities of the solution for the backprojection operator in tomography. *Proceedings of the American Mathematical Society*, 124(2): 567–577, 1996.

Alexander G Ramm and Alexander I Katsevich. *The Radon transform and local tomography.* CRC Press, 2020.

Michael Reed and Barry Simon. *Methods of Modern Mathematical Physics. Vol. 1: Functional Analysis.* Academic Press, San Diego, 1980.

Boris Rubin. The Calderón reproducing formula, windowed x-ray transforms, and Radon transforms in L^p-spaces. *Journal of Fourier Analysis and Applications*, 4(2):175–197, 1998.

Walter Rudin. *Functional Analysis.* McGraw-Hill, New York, 1973. McGraw-Hill Series in Higher Mathematics.

Walter Rudin. *Functional Analysis.* McGraw-Hill, New York, 2nd edition, 1991. McGraw-Hill Series in Higher Mathematics.

Stefan G Samko. Denseness of the Lizorkin-type spaces Φ_V in $L_p(\mathbb{R}^n)$. *Mathematical Notes of the Academy of Sciences of the USSR*, 31(6):432–437, 1982.

Stefan G Samko, Anatoly A Kilbas, and Oleg I Marichev. *Fractional Integrals and Derivatives: Theory and Applications*. Gordon and Breach Science Publishers, 1993.

Katerina Saneva and Jasson Vindas. Wavelet expansions and asymptotic behavior of distributions. *Journal of Mathematical Analysis and Applications*, 370(2):543–554, October 2010.

Bernhard Schölkopf, Kah-Kay Sung, Chris J C Burges, Federico Girosi, Partha Niyogi, Tomaso Poggio, and Vladimir Vapnik. Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *IEEE Transactions on Signal Processing*, 45 (11):2758–2765, November 1997.

Bernhard Schölkopf, Ralf Herbrich, and Alex J. Smola. A generalized representer theorem. In David Helmbold and Bob Williamson, editors, *Computational Learning Theory*, pages 416–426. Springer Berlin Heidelberg, 2001.

Donald C Solmon. Asymptotic formulas for the dual Radon transform and applications. *Mathematische Zeitschrift*, 195(3):321–343, 1987.

Sho Sonoda and Noboru Murata. Neural network with unbounded activation functions is universal approximator. *Applied and Computational Harmonic Analysis*, 43(2):233–268, 2017.

Michael Unser. A unifying representer theorem for inverse problems and machine learning. *Foundations of Computational Mathematics*, 21(4):941–960, September 2021. doi: 10.1007/s10208-020-09472-x.

Michael Unser and Shayan Aziznejad. Convex optimization in sums of Banach spaces. *Applied and Computational Harmonic Analysis*, 56:1–25, January 2022.

Michael Unser, Julien Fageot, and John Paul Ward. Splines are universal solutions of linear inverse problems with generalized-TV regularization. *SIAM Review*, 59(4):769–793, December 2017.

Grace Wahba. *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics, Philadelphia, 1990.