Noname manuscript No.
(will be inserted by the editor)

From Kernel Methods to Neural Networks: A Unifying Variational Formulation

Michael Unser

the date of receipt and acceptance should be inserted later

Abstract The minimization of a data-fidelity term and an additive regularization functional gives rise to a powerful framework for supervised learning. In this paper, we present a unifying regularization functional that depends on an operator L and on a generic Radon-domain norm. We establish the existence of a minimizer and give the parametric form of the solution(s) under very mild assumptions. When the norm is Hilbertian, the proposed formulation yields a solution that involves radial-basis functions and is compatible with the classical methods of machine learning. By contrast, for the total-variation norm, the solution takes the form of a two-layer neural network with an activation function that is determined by the regularization operator. In particular, we retrieve the popular ReLU networks by letting the operator be the Laplacian. We also characterize the solution for the intermediate regularization norms $\|\cdot\| = \|\cdot\|_{L_p}$ with $p \in (1, 2]$. Our framework offers guarantees of universal approximation for a broad family of regularization operators or, equivalently, for a wide variety of shallow neural networks, including the cases (such as ReLU) where the activation function is increasing polynomially. It also explains the favorable role of bias and skip connections in neural architectures.

Mathematics Subject Classification (2010) 44A12 · 46N10 · 47A52 · 68T07

Keywords machine learning · convex optimization · regularization · representer theorem · kernel methods · neural networks · Banach space

Communicated by Carola-Bibiane Schönlieb

The research leading to these results has received funding from the European Research Council under Grant ERC-2020-AdG FunLearn-101020573.

Biomedical Imaging Group, École polytechnique fédérale de Lausanne (EPFL), Station 17, CH-1015, Lausanne, Switzerland (`michael.unser@epfl.ch`).

Address(es) of author(s) should be given

1 Introduction

Regularization theory constitutes a powerful framework for the derivation of algorithms for supervised learning [14,41,42]. Given a series of data points $(\boldsymbol{x}_m, y_m) \in \mathbb{R}^d \times \mathbb{R}$, $m = 1, \dots, M$, the basic problem (regression) is to find a mapping $f : \mathbb{R}^d \to \mathbb{R}$ such that $f(\boldsymbol{x}_m) \approx y_m$, without overfitting. The standard paradigm is to let f be the minimizer of a cost that consists of a data-fidelity term and an additive regularization functional [8]. The minimization proceeds over a prescribed class $\mathcal{H}$ of candidate functions. One usually distinguishes between the parametric approaches (e.g., neural networks), where $\mathcal{H} = \mathcal{H}_\Theta$ is a family of functions specified by a finite set of parameters $\theta \in \Theta$ (e.g., the weights of the network), and the nonparametric ones, where the properties of the solution are controlled by the regularization functional. The focus of this paper is on the nonparametric techniques. They rely on *functional optimization*, which means that the minimization proceeds over a space of functions rather than over a set of parameters. The regularization is usually chosen to be an increasing function of the norm associated with a particular Banach space, which results in a well-posed problem [10,9,56].

The functional-optimization point of view is often constructive, in that it suggests or supports explicit learning architectures. For instance, the choice of the Hilbertian regularization $R(f) = \|f\|^2_{\mathcal{H}}$ where $\mathcal{H}$ is a reproducing kernel Hilbert space (RKHS) results in a closed-form solution that is a linear combination of kernels positioned on the data [62,7]. In fact, the RKHS setting yields a generic class of estimators that is compatible with the classical kernel-based methods of machine learning, including support vector machines [62,41,50,49, 1,56]. Likewise, adaptive kernel methods are justifiable from the minimization of a generalized total-variation norm, which favors sparse representions [11, 12,3]. These latter results actually take their root in spline theory [18,28,60]. Similarly, it has been demonstrated that shallow ReLU networks are solutions of functional-optimization problems with an appropriate regularization. One way to achieve this is to start from an explicit parameterization of an infinite-width network [4] (the reverse engineering/synthesis approach). Another way is to consider a regularization operator that is matched to the neuronal activation with a L_1-type penalty[1]; for instance, a second-order derivative for $d = 1$ [48,36] or, more generally, the Radon-domain counterpart of the Laplace operator whose Green's function is precisely a ReLU ridge [35,37,57]. Similar optimality results can be stated within the framework of reproducing-kernel Banach spaces [6], which is a formal point of view that bridges the synthesis and analysis approach of [4] and [37], respectively. Also relevant to the discussion is a variational formulation that links the ridgelet transform to the training of shallow neural networks with weight-decay regularization [53].

The second important benefit of the functional-optimization approach is that it gives insight on the approximation capabilities (expressivity) of the re-

[1] The precise formulation involves the $\mathcal{M}$-norm (or total variation), which is the weak form of L_1 associated with space of bounded Radon measures. In our account, we take it as the default norm for the Lebesgue exponent $p = 1$, with a slight abuse of language.

sulting learning architectures. This information is encapsulated in the choice/definition of the native space $\mathcal{H}$ (typically, a Sobolev space), which goes hand-in-hand with the regularization functional. Roughly speaking, the native space $\mathcal{H}$ ought to be "large enough" to allow for the approximation of any continuous function with an arbitrary degree of precision. This universal approximation property is a central theme in the theory of radial-basis functions (RBFs) [31, 63]. In machine learning, the kernel estimators that meet this approximation requirement are called *universal* [32]. When the basis functions are shifted replicates of a single template $h : \mathbb{R}^d \to \mathbb{R}$, then the condition is equivalent to h being strictly positive definite, which means that its Fourier transform is real-valued symmetric, and (strictly) positive [13]. Similar guarantees of universal approximation exist for (shallow) neural networks under mild conditions on the activation functions [16, 25, 30, 5, 39]. The main difference with the RKHS framework, however, is that the universality results for neural nets usually make the assumption that the input domain is a compact subset of $\mathbb{R}^d$.

The purpose of this paper is to unify and extend these various approaches by introducing a universal regularization functional. The latter has two components: an admissible differential operator L, and an L_p-type Radon-domain norm. The resulting regularization operator is $\mathrm{L_R} = \mathrm{K_{rad}RL}$, where R is the Radon transform and $\mathrm{K_{rad}}$ the "filtering" operator of computer tomography [33]. Our main result (Theorem 5) gives the parametric form of the solution of the corresponding functional-optimization problems under minimal hypotheses. For $p = 2$, the outcome of our theorem is compatible with the type of kernel expansions (RBFs) of classical machine learning for which there is a vast literature [52, 24]. For $p = 1$, the solution set is parameterized by a neural network with one hidden layer whose activation function is determined by the regularization operator. In particular, if we take L to be the Laplacian, then one retrieves the popular ReLU activation. Remarkably, the connection with neural networks also works the other way round: Parhi et al. [36, 38] could prove that the training of a shallow ReLU neural network that is sufficiently wide, with weight-decay regularization, converges to the solution of a functional optimization problem that is a special instance of the class considered in this paper.

The foundation for our characterization is an abstract representer theorem for direct-sum Banach spaces [58]. Thus, the primary effort in this paper consists in the development of a functional framework that is adapted to the Radon transform and that fulfills the hypotheses of the abstract theorem. The main contributions can be summarized as follows.

1. Construction and characterization of an extended family of native Banach spaces $\mathcal{X}'_{\mathrm{L_R}}(\mathbb{R}^d)$ associated with a generic Radon-domain norm $\|\cdot\|_{\mathcal{X}'}$ and a differential operator L, under the general admissibility conditions stated in Definition 3 (Theorem 6).
2. Proof that (i) the sampling functionals $\delta(\cdot - \boldsymbol{x}_m) : \mathcal{X}'_{\mathrm{L_R}}(\mathbb{R}^d) \to \mathbb{R}$ are weak*-continuous; and (ii) the adjoint of the regularization operator has a stable

generalized inverse $\mathrm{L}_{\mathrm{R}}^{*\dagger}$ (see Theorem 7 and accompanying explanations). These technical points are essential to the argumentation (existence of solution).

3. Extension and unification of a number of earlier optimality results for RBF expansions and neural networks. While the present setup for $p = 2$ and $\mathrm{L} = (-\Delta)^{\gamma}$ is reminiscent of thin-plate splines [17,29], the resulting solution for a fixed γ does not depend on the dimension d, which makes it easier to deploy. Likewise, our variational formulation with $\mathcal{X}' = \mathcal{M}$ extends the results of Parhi and Nowak [37] by: (i) proving that the neural network parameterization applies to *all the extreme points* of the solution set, and (ii) by covering a much broader class of activation functions, including those with polynomial growth (of degree n_0).
4. General guarantees of universality, subject to the admissibility condition in Definition 3. While the result for $p = 2$ is consistent with the known criteria for kernel estimators [32], its counterpart for neural networks ($\mathcal{X}' = \mathcal{M}$) brings in a new twist: the addition of a polynomial component. The latter, which is not present in the traditional theory [5,39], is necessary to lift the hypothesis of a compact input domain. The two cases of greatest practical relevance are the sigmoid and the ReLU activations which, in our formulation, require the addition of a bias ($n_0 = 0$) and an affine term ($n_0 = 1$), respectively. Interestingly, the ReLU case yields a neural architecture with a skip connection akin to ResNet [22], which is highly popular in practice.

The paper is organised as follows: We start with mathematical preliminaries in Section 2. In particular, we state our criteria of admissibility for L and show how to represent its polynomial null space. In Section 3, we review the main properties of the Radon transform and specify the dual pair $(\mathcal{X}_{\mathrm{Rad}}, \mathcal{X}'_{\mathrm{Rad}})$ of hyper-spherical Banach spaces that enter the definition of our native spaces. We also provide formulas for the (filtered) Radon transform of RBFs and ridges (the elementary constituents of neural networks). Section 4 is devoted to the description and interpretation of our main result (Theorem 5). In particular, we draw a connection with RKHS in Section 4.2. We discuss the issue of universality in Section 4.3 and show in Section 4.4 how our framework can be extended to handle antisymmetric activations, including sigmoids. We complement our exposition in Section 4.5 with a listing of specific configurations, many of which are intimately connected to splines. The mathematical developments that support our formulation are presented in Section 5. They include the characterization of the kernel of the inverse operator $\mathrm{L}_{\mathrm{R}}^{*\dagger}$—the enabling ingredient of our formulation—and the construction of the predual Banach space $\mathcal{X}_{\mathrm{L_R}}(\mathbb{R}^d)$.

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 $\boldsymbol{k} = (k_1, \dots, k_d) \in \mathbb{N}^d$ (where $\mathbb{N}$ includes 0) with the notational conventions $\boldsymbol{k}! = \prod_{i=1}^d k_i!$, $|\boldsymbol{k}| = k_1 + \cdots + k_d$, $\boldsymbol{x}^{\boldsymbol{k}} = \prod_{i=1}^d x_i^{k_i}$ for any $\boldsymbol{x} \in \mathbb{R}^d$, and $\partial^{\boldsymbol{k}} f(\boldsymbol{x}) = \frac{\partial^{|\boldsymbol{k}|} f(x_1,\dots,x_d)}{\partial_{x_1}^{k_1} \cdots \partial_{x_d}^{k_d}}$. This allows us to write the multidimensional Taylor expansion around $\boldsymbol{x} = \boldsymbol{x}_0$ of an analytical function $f : \mathbb{R}^d \to \mathbb{R}$ explicitly as

$$f(\boldsymbol{x}) = \sum_{n=0}^{\infty} \sum_{|\boldsymbol{k}|=n} \frac{\partial^{\boldsymbol{k}} f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0)^{\boldsymbol{k}}}{\boldsymbol{k}!} \tag{1}$$

where the internal summation is over all multi-indices $\boldsymbol{k}$ 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 $\|\cdot\|_{L_p}$, denoted as $L_p(\mathbb{R}^d) = \overline{(\mathcal{S}(\mathbb{R}^d), \|\cdot\|_{L_p})}$. For the endpoint $p = \infty$, we have $\overline{(\mathcal{S}(\mathbb{R}^d), \|\cdot\|_{L_\infty})} = C_0(\mathbb{R}^d)$ with $\|\varphi\|_{L_\infty} = \sup_{\boldsymbol{x} \in \mathbb{R}^d} |\varphi(\boldsymbol{x})|$, which is the space of continuous functions that vanish at infinity. The continuous dual of $C_0(\mathbb{R}^d)$ is the space $\mathcal{M}(\mathbb{R}^d) = \{f \in \mathcal{S}'(\mathbb{R}^d) : \|f\|_{\mathcal{M}} < \infty\}$ of bounded Radon measures with

$$\|f\|_{\mathcal{M}} = \sup_{\varphi \in \mathcal{S}(\mathbb{R}^d) : \|\varphi\|_{L_\infty} \le 1} \langle f, \varphi \rangle. \tag{2}$$

The latter is a superset of $L_1(\mathbb{R}^d)$, which is isometrically embedded in it, in the sense 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

$$\widehat{\varphi}(\boldsymbol{\omega}) \triangleq \mathcal{F}\{\varphi\}(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi(\boldsymbol{x}) \mathrm{e}^{-\mathrm{i}\langle \boldsymbol{\omega}, \boldsymbol{x} \rangle} \mathrm{d}\boldsymbol{x}. \tag{3}$$

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 (3).

To control the minimal order $\alpha \ge 0$ of decay (resp., the maximal rate of growth) of functions, we use the dual pair of spaces $L_{1,\alpha}(\mathbb{R}^d)$ and $L_{\infty,-\alpha}(\mathbb{R}^d) = \big(L_{1,\alpha}(\mathbb{R}^d)\big)'$. These are the Banach spaces associated with the weighted norms

$$\|f\|_{L_{1,\alpha}} \triangleq \int_{\mathbb{R}^d} (1 + \|\boldsymbol{x}\|)^\alpha |f(\boldsymbol{x})| \mathrm{d}\boldsymbol{x} \tag{4}$$

$$\|f\|_{L_{\infty,-\alpha}} \stackrel{\Delta}{=} \operatorname*{ess\,sup}_{\boldsymbol{x}\in\mathbb{R}} (1+\|\boldsymbol{x}\|)^{-\alpha}|f(\boldsymbol{x})|, \tag{5}$$

respectively. Specifically, the inclusion $f \in L_{\infty,-n_0}(\mathbb{R}^d)$ with $n_0 \in \mathbb{N}$ indicates that f cannot grow faster than a polynomial of degree n_0, while the condition $f \in L_{1,\alpha}(\mathbb{R}^d)$ implies that $f(\boldsymbol{x})$ must be locally integrable and must decay (slightly) faster than $1/\|\boldsymbol{x}\|^{\alpha+d}$ as $\|\boldsymbol{x}\| \to \infty$.

2.2 Admissible Regularization Operators

The regularization operators L that are of interest to us are linear, shift-invariant (LSI), and isotropic. For simplicity, we shall first specify the action of L on test functions, with the understanding that the domain of the operator will be extended to some corresponding "native space" that will be identified as we progress through the paper.

Definition 1 The linear operator $\mathrm{L} : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ is said to be

- *Shift-invariant*: If $\mathrm{L}\{\varphi(\cdot - \boldsymbol{x}_0)\} = \mathrm{L}\{\varphi\}(\cdot - \boldsymbol{x}_0)$ for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$ and $\boldsymbol{x}_0 \in \mathbb{R}^d$.
- *Isotropic (or rotation-invariant)*: If $\mathrm{L}\{\varphi(\mathbf{R}_{\boldsymbol{\theta}}\cdot)\} = \mathrm{L}\{\varphi\}(\mathbf{R}_{\boldsymbol{\theta}}\cdot)$ for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$ and any rotation matrix $\mathbf{R}_{\boldsymbol{\theta}}$ on $\mathbb{R}^d$.
- *Self-adjoint*: If the adjoint operator $\mathrm{L}^* : \mathcal{S}''(\mathbb{R}^d) = \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ has the same Schwartz kernel (impulse response) as L.

Since the Schwartz kernel of a linear operator is unique [55], the property of self-adjointness will be denoted as $\mathrm{L} = \mathrm{L}^*$, irrespective of the actual domain and range of the operator. It is well-known that a LSI operator can always be expressed as the convolution $\mathrm{L}\{\varphi\} = h * \varphi$, where $h = \mathrm{L}\{\delta\} \in \mathcal{S}'(\mathbb{R}^d)$ is the impulse response of L. When L is isotropic, h is a purely radial function. Since all isotropic functions are symmetric, this also implies that an isotropic LSI operator is necessarily self-adjoint. All such operators are characterized by a Fourier symbol (a.k.a. frequency response) $\widehat{L} = \mathcal{F}\{h\}$ that is purely radial, with $\widehat{L}(\boldsymbol{\omega}) = \widehat{L}_{\mathrm{rad}}(\|\boldsymbol{\omega}\|)$, under the implicit assumption that the radial profile $\widehat{L}_{\mathrm{rad}}$ is identifiable as a measurable function $\mathbb{R} \to \mathbb{R}$.

Our condition for admissibility is that L be invertible in an appropriate sense.

Definition 2 (Spline-admissible operators with trivial null space) An isotropic LSI operator L has a trivial null space if its radial frequency profile $\widehat{L}_{\mathrm{rad}}$ does not vanish over $\mathbb{R}$. We then say that it is *spline-admissible* if $1/\widehat{L}_{\mathrm{rad}} \in L_1(\mathbb{R})$ and $\rho_{\mathrm{rad}} = \mathcal{F}^{-1}\{1/\widehat{L}_{\mathrm{rad}}\} \in L_1(\mathbb{R})$ where the operator $\mathcal{F}^{-1} : L_1(\mathbb{R}) \to C_0(\mathbb{R})$ is the classical inverse Fourier transform.

The typical scenario is $\widehat{L}(\boldsymbol{\omega}) = (1 + \|\boldsymbol{\omega}\|^2)^{\alpha/2}$ with $\alpha \geq 1$, which results in a stable inverse operator L^{-1} whose radially symmetric impulse response is the Bessel potential of order α. These operators play a central in the theory of Sobolev spaces [21].

Distribution theory allows us to go further and to invert operators with nontrivial null spaces, but only if the zeros of the frequency response are located at isolated points. When the operator is isotropic, this reduces the options to the cases where $\widehat{L}(\boldsymbol{\omega})$ has a (multiple) zero at $\boldsymbol{\omega} = \mathbf{0}$. Specifically, we shall say that L is of order γ_0 if $|\widehat{L}(\boldsymbol{\omega})|/\|\boldsymbol{\omega}\|^{\gamma_0} = C_0$ as $\|\boldsymbol{\omega}\| \to 0$. The second important parameter is the asymptotic growth exponent of $\widehat{L}(\boldsymbol{\omega})$. This is the largest index γ_1 such that $|\widehat{L}(\boldsymbol{\omega})| \geq C_1\|\boldsymbol{\omega}\|^{\gamma_1}$, for all $\|\boldsymbol{\omega}\| > R$. It determines the smoothness of the Green's function of the operator.

Definition 3 (Spline-admissible operators with nontrivial null space) An isotropic LSI operator L with radial frequency profile $\widehat{L}_{\rm rad}$ is said to be *spline-admissible* with a polynomial null space of degree n_0 if the following conditions are satisfied.

1. The profile $\widehat{L}_{\rm rad}$ does not vanish over $\mathbb{R}$, except for a zero of order $\gamma_0 \in (n_0, n_0+1]$ at the origin; that is, $|\widehat{L}_{\rm rad}(\omega)|/|\omega|^{\gamma_0} = C_0$ as $\omega \to 0$.
2. There exists an order $\gamma_1 > 1$, a constant $C_1 > 0$, and a radius $R_1 > 0$ such that $|\widehat{L}_{\rm rad}(\omega)| \geq C_1|\omega|^{\gamma_1}$ for all $|\omega| > R_1$ (ellipticity).
3. For all $\varphi \in \mathcal{S}(\mathbb{R}^d)$, $\mathrm{L}^*\{\varphi\} \in L_{1,n_0}(\mathbb{R}^d)$.

The connection between Condition 1 and the null space of L will be explained in Section 2.3. Conditions 1 and 2 with $\gamma_1 > 1$ ensure that $\rho_{\rm rad} = \mathcal{F}^{-1}\{1/\widehat{L}_{\rm rad}\}$, which is the generalized inverse Fourier transform of the distribution $1/\widehat{L}_{\rm rad}$, is identifiable as a continuous function $\mathbb{R} \to \mathbb{R}$. The order γ_1 actually controls the degree of differentiability (Sobolev smoothness) of $\rho_{\rm rad}$. Condition 3 is a mild technical constraint on the decay of $\mathrm{L}^*\{\varphi\}$; this constraint has not appeared to be a practical limitation so far. For instance, if L is an ordinary differential operator (an arbitrary polynomial of the Laplace operator Δ) then $\mathrm{L}^*\{\varphi\} \in \mathcal{S}(\mathbb{R}^d)$, which is included in $L_{1,m}(\mathbb{R}^d)$ for any $m \in \mathbb{Z}$. We use this third condition for the handling of fractional operators whose impulse response decays slowly.

An attractive class of admissible operators with $\gamma_0 = \gamma_1 = \alpha$ and $n_0 = \lceil \alpha - 1 \rceil$ are the fractional Laplacians $(-\Delta)^{\frac{\alpha}{2}}$ with $\alpha \in (1, \infty)$ whose frequency response is $\|\boldsymbol{\omega}\|^\alpha$. The inverse of the fractional Laplacian of order α, which corresponds to a frequency-domain multiplication by $\|\boldsymbol{\omega}\|^{-\alpha}$, is denoted by $(-\Delta)^{-\frac{\alpha}{2}}$. Both operators are part of the same family (isotropic LSI and scale-invariant), their distributional impulse response being given by

$$k_{\alpha,d}(\boldsymbol{x}) = \mathcal{F}^{-1}\left\{\frac{1}{\|\boldsymbol{\omega}\|^\alpha}\right\}(\boldsymbol{x}) = \begin{cases} A_{\alpha,d}\,\|\boldsymbol{x}\|^{\alpha-d}, & \alpha - d, -\alpha \notin 2\mathbb{N} \\ B_{n,d}\,\|\boldsymbol{x}\|^{2n}\log(\|\boldsymbol{x}\|), & \alpha - d = 2n \in 2\mathbb{N} \\ (-\Delta)^n\{\delta\}, & -\alpha/2 = n \in \mathbb{N}, \end{cases} \tag{6}$$

with constants $A_{d,\alpha} = \frac{\Gamma\left(\frac{d-\alpha}{2}\right)}{2^\alpha \pi^{d/2}\Gamma\left(\frac{\alpha}{2}\right)}$ and $B_{d,n} = \frac{(-1)^{1+n}}{2^{2n+d-1}\pi^{d/2}\Gamma\left(n+\frac{d}{2}\right)n!}$ [19,47].

The kernel $k_{\alpha,d}$ can also be interpreted as the Green's function of $(-\Delta)^{\frac{\alpha}{2}}$, with the corresponding radial profile in Definition 3 being $\rho_{\rm rad}(t) = k_{\alpha,1}(t)$. In view of (6), this means that $(-\Delta)^{\frac{\alpha}{2}}$ is admissible for $\alpha > 1$.

We note that the impulse response of the filtering operator K in Theorem 1 is proportional to $k_{-d+1,d}(\boldsymbol{x})$, which tells us that it decays asymptotically like $1/\|\boldsymbol{x}\|^{2d-1}$ when d is even, or is a power of the Laplacian (local operator) otherwise. Functionally, this means that $\mathrm{K}\big(\mathcal{S}(\mathbb{R}^d)\big) = \mathcal{S}(\mathbb{R}^d)$ for even dimensions, and $\mathrm{K}\big(\mathcal{S}(\mathbb{R}^d)\big) \subset L_{1,d-1-\epsilon}(\mathbb{R}^d)$ otherwise. Likewise, the impulse response of the fractional Laplacians (of non-even order) decays asymptotically like $1/\|\boldsymbol{x}\|^{\alpha+d}$, which implies that $(-\Delta)^{\frac{\alpha}{2}}\big(\mathcal{S}(\mathbb{R}^d)\big) \subset L_{1,\alpha-\epsilon}(\mathbb{R}^d)$ for arbitrarily small $\epsilon > 0$, so that the third condition in Definition 3 is met.

2.3 Nontrivial Null Space and Related Projectors

Let L be a LSI operator whose frequency response $\widehat{L}$ satisfies the conditions

$$\partial^{\boldsymbol{k}}\widehat{L}(\boldsymbol{0}) = 0, \text{ for all } \boldsymbol{k} \in \mathbb{N}^d \text{ with } |\boldsymbol{k}| \le n_0, \tag{7}$$

for some integer $n_0 \ge 0$. This flatness behavior at the origin implies that L has the capacity to annihilate all polynomials of degree n_0 by mapping them to zero (see [61, p. 131]). The explanation lies in the property that the Fourier transform of any polynomial is entirely concentrated at the origin. If, in addition, we impose that $\widehat{L}(\boldsymbol{\omega}) \neq 0$ for all $\boldsymbol{\omega} \in \mathbb{R}^d\backslash\{\boldsymbol{0}\}$, then we are also making sure that the null space of L is limited to polynomials.

Next, we recall that the directional derivative of a function along the direction $\boldsymbol{\xi} \in \mathbb{S}^{d-1}$ (i.e., $\boldsymbol{\xi} \in \mathbb{R}^d$ with $\|\boldsymbol{\xi}\| = 1$) is given by

$$\mathrm{D}_{\boldsymbol{\xi}} f = \boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{\nabla} f = \xi_1\partial^{\boldsymbol{e}_1} f + \cdots + \xi_d\partial^{\boldsymbol{e}_d} f. \tag{8}$$

The operator $\mathrm{D}_{\boldsymbol{\xi}}$ is LSI with frequency response $\widehat{\mathrm{D}_{\boldsymbol{\xi}}}(\boldsymbol{\omega}) = (\mathrm{i}\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{\omega})$. The nth iterate of $\mathrm{D}_{\boldsymbol{\xi}}$ yields the nth derivative along $\boldsymbol{\xi}$ whose explicit expression in terms of partial derivatives is

$$\mathrm{D}^n_{\boldsymbol{\xi}} f(\boldsymbol{x}) = \mathcal{F}^{-1}\{(\mathrm{i}\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{\omega})^n \hat{f}(\boldsymbol{\omega})\}(\boldsymbol{x}) = \sum_{|\boldsymbol{k}|=n} \frac{n!}{\boldsymbol{k}!}\boldsymbol{\xi}^{\boldsymbol{k}}\partial^{\boldsymbol{k}} f(\boldsymbol{x}), \tag{9}$$

where the right-hand side follows from the application of the multinomial expansion to $(\mathrm{i}\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{\omega})^n = (\xi_1\mathrm{i}\omega_1 + \cdots + \xi_d\mathrm{i}\omega_d)^n$.

For isotropic operators, the directional derivatives $\mathrm{D}^n_{\boldsymbol{\xi}}\widehat{L}(\boldsymbol{0})$ do not dependent on the direction $\boldsymbol{\xi}$ and coincide with the radial derivatives $\widehat{L}^{(n)}_{\rm rad}(0)$. In view of (9), (7) then has the radial equivalent

$$\widehat{L}^{(n)}_{\rm rad}(0) = \frac{\mathrm{d}^n\widehat{L}_{\rm rad}(0)}{\mathrm{d}\omega^n} = 0, \text{ for } n = 0, 1, \dots, n_0, \tag{10}$$

which is much simpler to test. It follows that an operator whose radial frequency profile is such that $|\widehat{L}_{\rm rad}(\omega)|/|\omega|^{\gamma_0} = C_0$ as $\omega \to 0$ will annihilate all polynomials up to degree $n_0 = \lceil\gamma_0 - 1\rceil$.

Consequently, the null space of a spline-admissible operator L of order γ_0 consists of the polynomials of degree $n_0 = (\gamma_0 - 1)$ when γ_0 is an integer and $n_0 = \lfloor \gamma_0 \rfloor$ otherwise when $\gamma_0 \notin \mathbb{N}$. We shall represent these polynomials by expanding them in the monomial/Taylor basis

$$m_{\boldsymbol{k}}(\boldsymbol{x}) = \frac{\boldsymbol{x}^{\boldsymbol{k}}}{\boldsymbol{k}!} \tag{11}$$

with $|\boldsymbol{k}| \leq n_0$. We also add a topological structure by equipping the space with the ℓ_2 norm of the Taylor coefficients, which results in the description

$$\mathcal{P}_{n_0} = \Big\{p_0 = \sum_{|\boldsymbol{k}|\leq n_0} b_{\boldsymbol{k}} m_{\boldsymbol{k}} : \|p_0\|_{\mathcal{P}} < \infty\Big\} \text{ with } \|p_0\|_{\mathcal{P}} \triangleq \|(b_{\boldsymbol{k}})_{|\boldsymbol{k}|\leq n_0}\|_2. \tag{12}$$

To avoid a notational overload, we shall often denote this null space by $\mathcal{P}$, with the convention that $\mathcal{P} = \mathcal{P}_{n_0} = \{0\}$ when $n_0 = \lceil \gamma_0 - 1 \rceil < 0$ (for the operators L whose null space is trivial). The important point here is that (12) specifies a finite-dimensional Banach subspace of $\mathcal{S}'(\mathbb{R}^d)$. Its continuous dual $\mathcal{P}'$ is finite-dimensional as well, although it is composed of "abstract" elements $p_0^* \in \mathcal{P}'$ that are, in fact, equivalence classes in $\mathcal{S}'(\mathbb{R}^d)$. Yet, it is possible to identify every dual element $p_0^* \in \mathcal{P}'$ as a true function by selecting a particular dual basis $\{m^*_{\boldsymbol{k}}\}_{|\boldsymbol{k}|\leq n_0}$ such that $\langle m^*_{\boldsymbol{k}}, m_{\boldsymbol{k}'}\rangle = \delta_{\boldsymbol{k}-\boldsymbol{k}'}$ (Kroneker delta). Our specific choice is

$$m^*_{\boldsymbol{k}} = (-1)^{|\boldsymbol{k}|}\partial^{\boldsymbol{k}}\kappa_{\mathrm{iso}} \in \mathcal{S}(\mathbb{R}^d) \tag{13}$$

with $\boldsymbol{k} \in \mathbb{N}^d$, where κ_{iso} is the isotropic function described in Lemma 1.

Lemma 1 (adapted from [57]) *There exists an isotropic window* $\kappa_{\mathrm{iso}} \in \mathcal{S}(\mathbb{R}^d)$ *such that*

$$\langle m_{\boldsymbol{k}}, (-1)^{|\boldsymbol{n}|}\partial^{\boldsymbol{n}}\kappa_{\mathrm{iso}}\rangle = \delta_{\boldsymbol{k}-\boldsymbol{n}} \tag{14}$$

for all $\boldsymbol{k}, \boldsymbol{n} \in \mathbb{N}^d$*, subject to the spectral constraints* $\widehat{\kappa}_{\mathrm{iso}}(\boldsymbol{\omega}) = 1$ *for* $\|\boldsymbol{\omega}\| < \frac{1}{2}$*,* $1 \geq \widehat{\kappa}_{\mathrm{iso}}(\boldsymbol{\omega}) \geq 0$ *for* $\frac{1}{2} < \|\boldsymbol{\omega}\| < 1$*, and* $\widehat{\kappa}_{\mathrm{iso}}(\boldsymbol{\omega}) = 0$ *for* $\|\boldsymbol{\omega}\| \geq 1$*.*

This allows us to describe the dual space explicitly as

$$\mathcal{P}' = \mathcal{P}'_{n_0} = \Big\{p^*_0 = \sum_{|\boldsymbol{k}|\leq n_0} b^*_{\boldsymbol{k}} m^*_{\boldsymbol{k}} : \|p^*_0\|_{\mathcal{P}'} < \infty\Big\} \text{ with } \|p^*_0\|_{\mathcal{P}'} \triangleq \|(b^*_{\boldsymbol{k}})\|_2 \tag{15}$$

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

$$\mathrm{Proj}_{\mathcal{P}}\{f\} = \sum_{|\boldsymbol{k}|\leq n_0} \langle f, m^*_{\boldsymbol{k}}\rangle\, m_{\boldsymbol{k}}, \tag{16}$$

which is well-defined for any $f \in \mathcal{S}'(\mathbb{R}^d)$ since $m^*_{\boldsymbol{k}} \in \mathcal{S}(\mathbb{R}^d)$. The "transpose" of this operator is

$$\mathrm{Proj}_{\mathcal{P}'}\{\nu\} = \sum_{|\boldsymbol{k}|\le n_0} \langle m_{\boldsymbol{k}}, \nu\rangle\, m^*_{\boldsymbol{k}}, \tag{17}$$

which returns the projection of ν onto $\mathcal{P}'_{n_0} \subseteq \mathcal{S}(\mathbb{R}^d)$ under the implicit assumption that ν has sufficient decay for $\nu \mapsto \langle m_{\boldsymbol{k}}, \nu\rangle$ to be well-defined—for instance, $\nu \in L_{1,n_0}(\mathbb{R}^d)$. Correspondingly, we also have that $\mathrm{Proj}_{\mathcal{P}'}\{\mathrm{L}^*\varphi\} = 0$ for all φ such that $\mathrm{L}^*\{\varphi\} \in L_{1,n_0}(\mathbb{R}^d)$ since $\langle m_{\boldsymbol{k}}, \mathrm{L}^*\varphi\rangle = \langle \mathrm{L}m_{\boldsymbol{k}}, \varphi\rangle = 0$ for $|\boldsymbol{k}| \le n_0$. The latter manipulation of the duality product is legitimate in reason of the inclusion $\mathcal{P}_{n_0} \subset L_{\infty,-n_0}(\mathbb{R}^d) = \big(L_{1,n_0}(\mathbb{R}^d)\big)'$.

Even though the null space of an admissible operator L may be non-trivial, its intersection with $\mathcal{S}(\mathbb{R}^d)$ is always $\{0\}$. This implies that $\mathrm{L}^* = \mathrm{L}$ is injective on $\mathcal{S}(\mathbb{R}^d)$ with $\mathrm{L}^{*-1}\mathrm{L}^*\{\varphi\} = \varphi$ for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$ where $\mathrm{L}^{*-1} = \mathrm{L}^{-1}$ is the LSI operator whose frequency response is $1/|\widehat{L}|$.

3 Radon Transform

The Radon transform extracts the integrals of a function on $\mathbb{R}^d$ 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 $\mathbb{R}^d$. The coordinates of a hyperplane associated with an offset $t \in \mathbb{R}$ and a normal vector $\boldsymbol{\xi} \in \mathbb{S}^{d-1}$ satisfy

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

Here, we first review the classical theory of the Radon transform [27], starting with the case of test functions (Section 3.1), and extending it by duality to tempered distributions (Section 3.2). Then, in Section 3.3, we specify the Radon transform and its inverse on an appropriate class of intermediate Banach spaces $\mathcal{Y}$ with $\mathcal{S}(\mathbb{R}^d) \overset{\mathrm{d.}}{\hookrightarrow} \mathcal{Y} \overset{\mathrm{d.}}{\hookrightarrow} \mathcal{S}'(\mathbb{R}^d)$ (Theorem 3). Finally, in Section 3.4, we provide the (filtered) Radon transforms of the dictionary elements—isotropic kernels and ridges—that are relevant to our investigation.

3.1 Classical Integral Formulation

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

$$\mathrm{R}\{f\}(t,\boldsymbol{\xi}) = \int_{\mathbb{R}^d} \delta(t - \boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}) f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}, \quad (t,\boldsymbol{\xi}) \in \mathbb{R} \times \mathbb{S}^{d-1}. \tag{18}$$

The adjoint of R is the backprojection operator R^*. Its action on $g : \mathbb{R}\times\mathbb{S}^{d-1} \to \mathbb{R}$ yields the function

$$\mathrm{R}^*\{g\}(\boldsymbol{x}) = \int_{\mathbb{S}^{d-1}} g(\underbrace{\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}}_{t}, \boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}, \quad \boldsymbol{x} \in \mathbb{R}^d. \tag{19}$$

Given the d-dimensional Fourier transform $\hat{f}$ of $f \in L_1(\mathbb{R}^d)$, one can calculate $\mathrm{R}\{f\}(\cdot, \boldsymbol{\xi}_0)$ for any fixed $\boldsymbol{\xi}_0 \in \mathbb{S}^{d-1}$ through the relation

$$\mathrm{R}\{f\}(t, \boldsymbol{\xi}_0) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(\omega \boldsymbol{\xi}_0) \mathrm{e}^{\mathrm{i}\omega t} \mathrm{d}\omega = \mathcal{F}_{\mathrm{1D}}^{-1}\{\hat{f}(\cdot \boldsymbol{\xi}_0)\}\{t\}, \tag{20}$$

In other words, the restriction of $\hat{f} : \mathbb{R}^d \to \mathbb{C}$ along the ray $\{\boldsymbol{\omega} = \omega \boldsymbol{\xi}_0 : \omega \in \mathbb{R}\}$ coincides with the 1D Fourier transform of $\mathrm{R}\{f\}(\cdot, \boldsymbol{\xi}_0)$, a property that is referred to as the *Fourier-slice theorem.*

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. There, the Euclidean indexing with $\boldsymbol{x} \in \mathbb{R}^d$ must be 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 ϕ is represented by the duality product $g : \phi \mapsto \langle g, \phi \rangle_{\mathrm{Rad}}$. When g can be identified as an ordinary function $g : (t, \boldsymbol{\xi}) \mapsto g(t, \boldsymbol{\xi}) \in \mathbb{R}$, one can write that

$$\langle g, \phi \rangle_{\mathrm{Rad}} = \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} g(t, \boldsymbol{\xi}) \phi(t, \boldsymbol{\xi}) \mathrm{d}t \mathrm{d}\boldsymbol{\xi} \tag{21}$$

where $\mathrm{d}\boldsymbol{\xi}$ stands for the surface element on the unit sphere $\mathbb{S}^{d-1}$.

The key property for analysis is that the Radon transform is continuous on $\mathcal{S}(\mathbb{R}^d)$ and invertible [27,23,43]. In addition to a backprojection, the inversion involves the so-called "filtering" operator.

Definition 4 The filtering operator $\mathrm{K} : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ is defined as

$$\mathrm{K}\{\varphi\} = \mathcal{F}^{-1}\{\widehat{K}\hat{\varphi}\} \quad \text{with} \quad \widehat{K}(\boldsymbol{\omega}) = c_d \|\boldsymbol{\omega}\|^{d-1}, \tag{22}$$

where $c_d = \frac{1}{2(2\pi)^{d-1}}$.

The filtering operator is isotropic LSI and, as such, has a Radon-domain counterpart (see Definition 5) denoted by $\mathrm{K}_{\mathrm{rad}}$ that exclusively acts along the radial variable.

Definition 5 (Radon-domain counterpart of an isotropic LSI operator) Let $\mathrm{L} : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ be an isotropic LSI operator with radial frequency profile $\widehat{L}_{\mathrm{rad}} : \mathbb{R} \to \mathbb{R}$. Then, the Radon-domain counterpart of L is $\mathrm{L}_{\mathrm{rad}} : \mathcal{S}(\mathbb{R} \times \mathbb{S}^{d-1}) \to \mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1})$, which is defined as

$$\mathrm{L}_{\mathrm{rad}}\{\phi(\cdot, \boldsymbol{\xi})\}(t) = \mathcal{F}_{\mathrm{1D}}^{-1}\{\widehat{L}_{\mathrm{rad}} \hat{\phi}(\cdot, \boldsymbol{\xi})\}(t), \tag{23}$$

where $\hat{\phi}(\omega, \boldsymbol{\xi}) = \int_{\mathbb{R}} \phi(t, \boldsymbol{\xi}) \mathrm{e}^{-\mathrm{i}\omega t} \mathrm{d}t$ is the 1D Fourier transform of $t \mapsto \phi(t, \boldsymbol{\xi})$.

Theorem 1 (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,* $\mathrm{K}\mathrm{R}^*\mathrm{R} = \mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\mathrm{R} = \mathrm{R}^*\mathrm{R}\mathrm{K} = \mathrm{Id}$ *on* $\mathcal{S}(\mathbb{R}^d)$.

Based on this result, we can identify the filtering operator as $\mathrm{K} = (\mathrm{R}^*\mathrm{R})^{-1} = c_d(-\Delta)^{\frac{d-1}{2}}$ (fractional Laplacian). Alternatively, one can perform the filtering in the Radon domain by means of the operator $\mathrm{K}_{\mathrm{rad}}$, which implements a 1D convolution along the radial variable. The connection is that the frequency response of $\mathrm{K}_{\mathrm{rad}}$ coincides with the radial frequency profile of K so that $\widehat{K}(\boldsymbol{\omega}) = \widehat{K}_{\mathrm{rad}}(\|\boldsymbol{\omega}\|)$ with $\widehat{K}_{\mathrm{rad}}(\omega) = c_d|\omega|^{d-1}$.

While the Radon transform $\mathrm{R} : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R} \times \mathbb{S}^{d-1})$ is invertible, it is not surjective, which means that not every hyper-spherical test function $\phi \in \mathcal{S}(\mathbb{R} \times \mathbb{S}^{d-1})$ can be written as $\phi = \mathrm{R}\{\varphi\}$ with $\varphi \in \mathcal{S}(\mathbb{R}^d)$. A necessary condition is that ϕ be even, but this is not sufficient [20,27,23]. The good news, however, is that the range of R on $\mathcal{S}(\mathbb{R}^d)$ is a closed subspace of $\mathcal{S}(\mathbb{R} \times \mathbb{S}^{d-1})$ [23, p. 60]. Accordingly, one can identify the range space $\mathcal{S}_{\mathrm{Rad}} \triangleq \mathrm{R}\big(\mathcal{S}(\mathbb{R}^d)\big)$ equipped with the Fréchet topology of $\mathcal{S}(\mathbb{R} \times \mathbb{S}^{d-1})$. Since the domain and range spaces are both Fréchet, we then invoke the open-mapping theorem [45, Theorem 2.11] to deduce that the transform $\varphi \mapsto \mathrm{R}\{\varphi\}$ is a homeomorphism of $\mathcal{S}(\mathbb{R}^d)$ onto $\mathcal{S}_{\mathrm{Rad}}$.

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

3.2 Distributional Extension

To extend the framework to distributions, one proceeds by duality. By invoking the property that $\mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\mathrm{R} = \mathrm{Id}$ on $\mathcal{S}(\mathbb{R}^d)$, we make the manipulation

$$\begin{aligned}\forall \varphi \in \mathcal{S}(\mathbb{R}^d) \quad \langle f, \varphi\rangle &= \langle f, \mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{\varphi\}\rangle \\ &= \langle \mathrm{R}\{f\}, \mathrm{K}_{\mathrm{rad}}\mathrm{R}\{\varphi\}\rangle_{\mathrm{Rad}} = \langle \mathrm{R}\{f\}, \phi\rangle_{\mathrm{Rad}}, \end{aligned} \tag{24}$$

with $\phi = \mathrm{K}_{\mathrm{rad}}\mathrm{R}\{\varphi\} \in \mathrm{K}_{\mathrm{rad}}\mathrm{R}\big(\mathcal{S}(\mathbb{R}^d)\big)$ and $\varphi = \mathrm{R}^*\{\phi\}$. Relation (24), which is valid in the classical sense for $f \in L_1(\mathbb{R}^d)$, is then used as definition to extend the scope of R for $f \in \mathcal{S}'(\mathbb{R}^d)$.

Definition 6 The distribution $g = \mathrm{R}\{f\} \in \big(\mathrm{K}_{\mathrm{rad}}\mathrm{R}\big(\mathcal{S}(\mathbb{R}^d)\big)\big)'$ is the (formal) Radon transform of $f \in \mathcal{S}'(\mathbb{R}^d)$ if

$$\forall \phi \in \mathrm{K}_{\mathrm{rad}}\mathrm{R}\big(\mathcal{S}(\mathbb{R}^d)\big) : \quad \langle g, \phi\rangle_{\mathrm{Rad}} = \langle f, \mathrm{R}^*\{\phi\}\rangle. \tag{25}$$

Likewise, $\tilde{g} = \mathrm{K}_{\mathrm{rad}}\mathrm{R}\{f\} \in \mathcal{S}'_{\mathrm{Rad}}$ is the (formal) filtered projection of $f \in \mathcal{S}'(\mathbb{R}^d)$ if

$$\forall \phi \in \mathcal{S}_{\mathrm{Rad}} : \quad \langle \tilde{g}, \phi\rangle_{\mathrm{Rad}} = \langle f, \mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\{\phi\}\rangle. \tag{26}$$

Finally, $f = \mathrm{R}^*\{g\} \in \mathcal{S}'(\mathbb{R}^d)$ is the backprojection of $g \in \mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1})$ if

$$\forall \varphi \in \mathcal{S}(\mathbb{R}^d) : \quad \langle \mathrm{R}^*\{g\}, \varphi\rangle = \langle g, \mathrm{R}\{\varphi\}\rangle_{\mathrm{Rad}}. \tag{27}$$

While (27) identifies $\mathrm{R}^*\{g\}$ as a single, unique distribution in $\mathcal{S}'(\mathbb{R}^d)$, this is not so for (26) (resp., (25)), as the members of $\mathcal{S}'_{\mathrm{Rad}}$ (resp., of $\big(\mathrm{K}_{\mathrm{rad}}\mathrm{R}\big(\mathcal{S}\big)\big)'$) are equivalence classes in $\mathcal{S}'(\mathbb{R}\times\mathbb{S}^{d-1})$. To make this explicit, we take advantage of the equivalence $\mathrm{R}^*\{g\}=0 \Leftrightarrow \langle g,\phi\rangle_{\mathrm{Rad}}=0$ to identity the null space of the backprojection operator as being

$$\mathcal{N}_{\mathrm{R}^*}=\{g\in\mathcal{S}'(\mathbb{R}\times\mathbb{S}^{d-1}):\langle g,\phi\rangle_{\mathrm{Rad}}=0,\forall\phi\in\mathcal{S}_{\mathrm{Rad}}\},\tag{28}$$

which is a closed subspace of $\mathcal{S}'(\mathbb{R}\times\mathbb{S}^{d-1})$. It is then possible to describe $\mathcal{S}'_{\mathrm{Rad}}$ as the abstract quotient space $\mathcal{S}'(\mathbb{R}\times\mathbb{S}^{d-1})/\mathcal{N}_{\mathrm{R}^*}$. In other words, if we find a hyper-spherical distribution $g_0\in\mathcal{S}'(\mathbb{R}\times\mathbb{S}^{d-1})$ such that (26) is met for a given $f\in\mathcal{S}'(\mathbb{R}^d)$, then, strictly speaking, $\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{f\}\in\mathcal{S}'_{\mathrm{Rad}}$ is the equivalence class (or coset) given by

$$\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{f\}=[g_0]=\{g_0+h:h\in\mathcal{N}_{\mathrm{R}^*}\}.\tag{29}$$

Since $[g_0]=[g]$ for any $g\in\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{f\}$, we refer to the members of $\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{f\}$ as "formal" filtered projections of f to remind us of this lack of unicity.

Based on those definitions, one obtains the classical result on the invertibility of the (filtered) Radon transform on $\mathcal{S}'(\mathbb{R}^d)$ [27], which is the dual of Corollary 1.

Theorem 2 (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* $\mathcal{S}'(\mathbb{R}^d)$. *More precisely, the filtered Radon transform* $\mathrm{K}_{\mathrm{rad}}\mathrm{R}:\mathcal{S}'(\mathbb{R}^d)\to\mathcal{S}'_{\mathrm{Rad}}$ *is a continuous bijection, with a continuous inverse given by* $(\mathrm{K}_{\mathrm{rad}}\mathrm{R})^{-1}=\mathrm{R}^*:\mathcal{S}'_{\mathrm{Rad}}\to\mathcal{S}'(\mathbb{R}^d)$.

To illustrate the fact that (26) does not identify a single distribution, we consider the Dirac ridge $\delta(\boldsymbol{\xi}_0\boldsymbol{x}-t_0)\in\mathcal{S}'(\mathbb{R}^d)$ and refer to the definition (18) of the Radon transform to deduce that, for all $\phi=\mathrm{R}\{\varphi\}\in\mathcal{S}_{\mathrm{Rad}}$ with $\varphi\in\mathcal{S}(\mathbb{R}^d)$,

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

which shows that $\delta\big(\cdot+\boldsymbol{z}_0\big)$ with $\boldsymbol{z}_0=(t_0,\boldsymbol{\xi}_0)$ 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 $\delta(\cdot-\boldsymbol{z}_0)$, as well as for $\delta_{\mathrm{Rad},\boldsymbol{z}_0}\stackrel{\Delta}{=}\frac{1}{2}\big(\delta(\cdot-\boldsymbol{z}_0)+\delta(\cdot+\boldsymbol{z}_0)\big)$, which has the advantage of being symmetric. While the general solution in $\mathcal{S}'_{\mathrm{Rad}}$ is $\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{\delta(\boldsymbol{\xi}_0^{\mathsf{T}}\cdot-t_0)\}=[\delta\big(\cdot\pm\boldsymbol{z}_0\big)]$, we shall see that there also exists a way to specify a representer that is unique (i.e., $\delta_{\mathrm{Rad},\boldsymbol{z}_0}$) by restricting the range of $\mathrm{K}_{\mathrm{rad}}\mathrm{R}$ to a suitable subspace of measures.

The distributional extension of the Radon transform inherits most of the properties of the "classical" operator defined in (18). 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,

$$\mathrm{R}\{h * f\} = \mathrm{R}\{h\} \circledast \mathrm{R}\{f\} \tag{30}$$

where the symbol "$\circledast$" denotes the 1D convolution along the radial variable $t \in \mathbb{R}$ with $(u \circledast g)(t, \boldsymbol{\xi}) = \langle u(\cdot, \boldsymbol{\xi}), g(t - \cdot, \boldsymbol{\xi})\rangle$. In particular, when $h = \mathrm{L}\{\delta\}$ is the (isotropic) impulse response of an LSI operator whose frequency response $\widehat{L}(\boldsymbol{\omega}) = \widehat{L}_{\mathrm{rad}}(\|\boldsymbol{\omega}\|)$ is purely radial, we get that

$$\mathrm{R}\{h * f\} = \mathrm{RL}\{f\} = \mathrm{L}_{\mathrm{rad}}\mathrm{R}\{f\}, \tag{31}$$

where $\mathrm{L}_{\mathrm{rad}}$ is the corresponding Radon-domain operator of Definition 5. Likewise, by duality, for $g \in \mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1})$ we have that

$$\mathrm{LR}^*\{g\} = \mathrm{R}^*\mathrm{L}_{\mathrm{rad}}\{g\}, \tag{32}$$

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 1, we can then use these relations for $\mathrm{L} = \mathrm{K} = (\mathrm{R}^*\mathrm{R})^{-1}$ to show that $\mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{f\} = \mathrm{R}^*\mathrm{RK}\{f\} = \mathrm{KR}^*\mathrm{R}\{f\} = f$ for a broad class of distributions. The first form is valid for all $f \in \mathcal{S}'(\mathbb{R}^d)$ (Theorem 2), but there is a slight restriction with the second form (resp., third form), which requires that $\mathrm{K}\{f\}$ $\big($resp., $\mathrm{K}\{g\}$ with $g = \mathrm{R}^*\mathrm{R}\{f\} \in \mathcal{S}'(\mathbb{R}^d)\big)$ be well-defined in $\mathcal{S}'(\mathbb{R}^d)$. While the latter condition is always met when d is odd, it may fail[2] in even dimensions with distributions (e.g., polynomials) whose Fourier transform is singular at the origin. The good news for our regularization framework is that these problematic distributions are either excluded from the native space or annihilated by L, so that it is legitimate to write that $\mathrm{L}_{\mathrm{R}} = \mathrm{K}_{\mathrm{rad}}\mathrm{RL} = \mathrm{RKL}$, where the second form has the advantage that K and L can be pooled into a single augmented operator (KL). An alternative form is $\mathrm{L}_{\mathrm{R}} = \mathrm{Q}_{\mathrm{rad}}\mathrm{R}$, where $\mathrm{Q}_{\mathrm{rad}} = \mathrm{K}_{\mathrm{rad}}\mathrm{L}_{\mathrm{rad}}$ is the radial Radon-domain operator whose frequency response is $\widehat{Q}_{\mathrm{rad}}(\omega) = c_d|\omega|^{d-1}\widehat{L}_{\mathrm{rad}}(\omega)$.

3.3 Radon-Compatible Banach Spaces

Our formulation requires the identification of Radon-domain Banach spaces over which the backprojection operator R^* is invertible. This is a non-trivial point because the extended operator $\mathrm{R}^* : \mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1}) \to \mathcal{S}'(\mathbb{R}^d)$ in Definition 6 is not injective. In fact, it has the highly non-trivial null space $\ker(\mathrm{R}^*) = \mathcal{S}_{\mathrm{Rad}}^{\perp}$, which is a superset of the odd Radon-domain distributions [20]. Yet, R^* is invertible on $\mathcal{S}'_{\mathrm{Rad}}$ and surjective on $\mathcal{S}'(\mathbb{R}^d)$ (Theorem 2).

To ensure invertibility, we therefore need to restrict ourselves to Banach spaces that are embedded in $\mathcal{S}'_{\mathrm{Rad}}$. To identify such objects, we consider a

[2] 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}. This means that K can properly handle (and annihilate) polynomials only up to degree $(2n-2)$.

generic 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 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 right-hand 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 $\|g\|_{\mathcal{X}'}$ given by the right-hand 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
$$\mathcal{X}' = \big\{g \in \mathcal{S}'(\mathbb{R}\times\mathbb{S}^{d-1}) : \|g\|_{\mathcal{X}'} < \infty\big\}. \tag{35}$$

Likewise, based on the pair $(\mathcal{S}_{\mathrm{Rad}}, \mathcal{S}'_{\mathrm{Rad}})$, we specify the 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 $\mathcal{S}_{\mathrm{Rad}}$ and $\mathcal{X}_{\mathrm{Rad}}$ substituting for $\mathcal{S}(\mathbb{R}\times\mathbb{S}^{d-1})$ and $\mathcal{X}$.

Theorem 3 (adapted from [57]) *Let $(\mathcal{X}_{\mathrm{Rad}}, \mathcal{X}'_{\mathrm{Rad}})$ be the dual pair of spaces specified by* (36) *and* (37)*. Then,*

1. *the map $\mathrm{R}^*\mathrm{K}_{\mathrm{rad}} : \mathcal{X}_{\mathrm{Rad}} \to \mathcal{Y} = \mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\big(\mathcal{X}_{\mathrm{Rad}}\big)$ is an isometric bijection, with $\mathrm{R}\mathrm{R}^*\mathrm{K}_{\mathrm{rad}} = \mathrm{Id}$ on $\mathcal{X}_{\mathrm{Rad}}$;*
2. *the map $\mathrm{R}^* : \mathcal{X}'_{\mathrm{Rad}} \to \mathcal{Y}' = \mathrm{R}^*\big(\mathcal{X}'_{\mathrm{Rad}}\big)$ is an isometric bijection, with $\mathrm{K}_{\mathrm{rad}}\mathrm{R}\mathrm{R}^* = \mathrm{Id}$ on $\mathcal{X}'_{\mathrm{Rad}}$.*

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 $\mathcal{X}' = \mathcal{X}'_{\mathrm{Rad}} \oplus (\mathcal{X}^{\mathrm{c}}_{\mathrm{Rad}})'$ where $(\mathcal{X}^{\mathrm{c}}_{\mathrm{Rad}})'$ can be identified as the null space of the backprojection operator $\mathrm{R}^* : \mathcal{X}' \to \mathcal{Y}' \hookrightarrow \mathcal{S}(\mathbb{R}^d)$.

The 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 $(\mathcal{X}, \mathcal{X}') = \big(C_0(\mathbb{R}\times\mathbb{S}^{d-1}), \mathcal{M}(\mathbb{R}\times\mathbb{S}^{d-1})\big)$. In fact, those hyper-spherical spaces have the convenient feature of admitting a decomposition in their even and odd components.

Lemma 2 *Let $\mathcal{Z} = \mathbb{R}\times\mathbb{S}^{d-1}$. Then, for $\mathcal{X} = L_p(\mathcal{Z})$ with $p \in [1,\infty)$ and $\mathcal{X} = C_0(\mathcal{Z})$ for $p = \infty$, we have that $\mathcal{X} = \mathcal{X}_{\mathrm{Rad}} \oplus \mathcal{X}^{\mathrm{c}}_{\mathrm{Rad}}$ where*
$$\mathcal{X}_{\mathrm{Rad}} = \mathcal{X}_{\mathrm{even}} = \{g \in \mathcal{X} : g(\boldsymbol{z}) = g(-\boldsymbol{z}), \forall \boldsymbol{z} \in \mathcal{Z}\} \tag{38}$$
$$\mathcal{X}^{\mathrm{c}}_{\mathrm{Rad}} = \mathcal{X}_{\mathrm{odd}} = \{g \in \mathcal{X} : g(\boldsymbol{z}) = -g(-\boldsymbol{z}), \forall \boldsymbol{z} \in \mathcal{Z}\}. \tag{39}$$

Proof To establish this result directly is tricky because the characterization of $\mathcal{S}_{\mathrm{Rad}}$ involves some general moment conditions [20, 23, 27]. Instead, we consider the smaller space of even Radon-domain Lizorkin test functions [26] described by

$$\mathcal{S}_{\mathrm{Liz,Rad}} = \{\phi \in \mathcal{S}_{\mathrm{even}}(\mathcal{Z}) : \int_{\mathbb{R}} t^k \phi(t, \boldsymbol{\xi})\mathrm{d}t = 0, \forall \boldsymbol{\xi} \in \mathbb{S}^{d-1}, k \in \mathbb{N}\}, \tag{40}$$

which is such that $\mathcal{S}_{\mathrm{Liz,Rad}} \subset \mathcal{S}_{\mathrm{Rad}} \subset \mathcal{S}_{\mathrm{even}}(\mathcal{Z})$. We then invoke a general result by Samko [46] that implies that $\overline{(\mathcal{S}_{\mathrm{Liz,Rad}}, \|\cdot\|_{L_p})} = L_{p,\mathrm{even}}(\mathcal{Z}) \supset (\mathcal{S}_{\mathrm{Rad}}, \|\cdot\|_{L_p})$ for $p \in [1, \infty)$ and $\overline{(\mathcal{S}_{\mathrm{Liz,Rad}}, \|\cdot\|_{L_\infty})} = C_{0,\mathrm{even}}(\mathcal{Z})$ otherwise [34]. The claim then follows from the observation that $L_p(\mathcal{Z}) = L_{p,\mathrm{even}}(\mathcal{Z}) \oplus L_{p,\mathrm{odd}}(\mathcal{Z})$ with $L_{p,\mathrm{even}}(\mathcal{Z}) = \overline{(\mathcal{S}_{\mathrm{Rad}}, \|\cdot\|_{L_p})}$ (because the completion is unique) and suitable adaptation for $p = \infty$. □

Correspondingly, we get that $\mathcal{X}'_{\mathrm{Rad}} = \mathrm{P}_{\mathrm{even}}(\mathcal{X}') = \mathcal{X}'_{\mathrm{even}}$ and $(\mathcal{X}^{\mathrm{c}}_{\mathrm{Rad}})' = (\mathrm{Id} - \mathrm{P}_{\mathrm{even}})(\mathcal{X}') = \mathcal{X}'_{\mathrm{odd}}$, with the cases of greatest interest to us being $\mathcal{M}_{\mathrm{Rad}} = \mathcal{M}_{\mathrm{even}}(\mathbb{R} \times \mathbb{S}^{d-1})$ and $L_{2,\mathrm{Rad}} = L_{2,\mathrm{even}}(\mathbb{R} \times \mathbb{S}^{d-1})$.

3.4 Specific Radon Transforms

The Fourier-slice theorem expressed by (20) remains valid for tempered distributions [43] and therefore also yields a characterization of $\mathrm{R}\{f\}$ that is compatible with the Banach framework of Theorem 3. It is especially helpful when the underlying function ρ_{iso} is isotropic with a known radial frequency profile $\widehat{\rho}_{\mathrm{rad}}$ such that $\mathcal{F}\{\rho_{\mathrm{iso}}\}(\boldsymbol{\omega}) = \widehat{\rho}_{\mathrm{rad}}(\|\boldsymbol{\omega}\|)$.

Proposition 1 (Radon transform of isotropic distributions) *Let ρ_{iso} be an isotropic distribution whose radial frequency profile is $\widehat{\rho}_{\mathrm{rad}} : \mathbb{R} \to \mathbb{R}$. Then,*

$$\mathrm{R}\{\rho_{\mathrm{iso}}(\cdot - \boldsymbol{x}_0)\}(t, \boldsymbol{\xi}) = \rho_{\mathrm{rad}}(t - \boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}_0) \tag{41}$$

$$\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{\rho_{\mathrm{iso}}(\cdot - \boldsymbol{x}_0)\}(t, \boldsymbol{\xi}) = \nu_{\mathrm{rad}}(t - \boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}_0) \tag{42}$$

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

The other important building blocks for representing functions are ridges. Specifically, a ridge is a multidimensional function

$$r_{\boldsymbol{\xi}_0} : \mathbb{R}^d \to \mathbb{R} : \boldsymbol{x} \mapsto r(\boldsymbol{\xi}_0^{\mathsf{T}}\boldsymbol{x}) \tag{43}$$

that is characterized by a profile $r : \mathbb{R} \to \mathbb{R}$ and a direction $\boldsymbol{\xi}_0 \in \mathbb{S}^{d-1}$. In effect, $r_{\boldsymbol{\xi}_0}$ varies along the axis specified by $\boldsymbol{\xi}_0$ and is constant within any hyperplane perpendicular to $\boldsymbol{\xi}_0$. The connection between ridges and the Radon transform is given by the *ridge identity*

$$\forall \varphi \in \mathcal{S}(\mathbb{R}^d) : \quad \langle r_{\boldsymbol{\xi}_0}, \varphi\rangle = \langle r, \mathrm{R}\{\varphi\}(\cdot, \boldsymbol{\xi}_0)\rangle, \tag{44}$$

where the right-hand side duality product (1D) is well-defined for any $r \in \mathcal{S}'(\mathbb{R})$ because $\mathrm{R}\{\varphi\}(\cdot, \boldsymbol{\xi}_0) \in \mathcal{S}(\mathbb{R})$ (by Theorem 1). When the profile $r : \mathbb{R} \to \mathbb{R}$ is locally integrable, (44) is a simple consequence of Fubini's theorem. For more general distributional profiles $r \in \mathcal{S}'(\mathbb{R})$, we use the ridge identity as definition, which then leads to the following characterization [57].

Theorem 4 (Filtered Radon transform of a ridge) *The filtered Radon transform of the (generalized) ridge $r_{\boldsymbol{\xi}_0}$ with profile $r \in \mathcal{S}'(\mathbb{R}^d)$ and direction $\boldsymbol{\xi}_0 \in \mathbb{S}^{d-1}$ is given by*

$$\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{r_{\boldsymbol{\xi}_0}\}(t, \boldsymbol{\xi}) = [r(t)\delta(\boldsymbol{\xi} - \boldsymbol{\xi}_0)], \tag{45}$$

where $[r(t)\delta(\boldsymbol{\xi} - \boldsymbol{\xi}_0)] \in \mathcal{S}'_{\mathrm{Rad}}$ is the equivalence class of distributions specified by (29). If $r \in \mathcal{M}(\mathbb{R})$, then the latter has the unique, concrete representer

$$\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{r_{\boldsymbol{\xi}_0}\}(t, \boldsymbol{\xi}) = \frac{1}{2}\big(r(t)\delta(\boldsymbol{\xi} - \boldsymbol{\xi}_0) + r(-t)\delta(\boldsymbol{\xi} + \boldsymbol{\xi}_0)\big) \tag{46}$$

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

An important special case of (46) is the Radon transform of a Dirac ridge: $\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{\delta(\boldsymbol{\xi}_0^{\mathsf{T}} \cdot - t_0)\} = \delta_{\mathrm{Rad},(t_0,\boldsymbol{\xi}_0)} = \frac{1}{2}\big(\delta(\cdot - t_0)\delta(\cdot - \boldsymbol{\xi}_0) + \delta(\cdot + t_0)\delta(\cdot + \boldsymbol{\xi}_0)\big)$, which has already been mentioned in Section 3.2 (see also [35, Example 1]).

4 Unifying Variational Formulation

4.1 Representer Theorem for Radon-Domain Regularization

From now on, we shall use the generic symbol $\mathcal{X}$ to designate the hyperspherical Banach space $L_q(\mathbb{R} \times \mathbb{S}^{d-1})$ with $q = (1, \infty)$ or $C_0(\mathbb{R} \times \mathbb{S}^{d-1})$ for $q = \infty$, which fall into the category described by (33) with $\|\cdot\|_{\mathcal{X}} = \|\cdot\|_{L_q}$.

The formulation of Theorem 5 below requires the specification of a *native space*, $\mathcal{X}'_{\mathrm{L_R}}(\mathbb{R}^d)$, that is tied to a Radon-domain norm $\|\cdot\|_{\mathcal{X}'}$ and an admissible regularization operator $\mathrm{L_R}$. Informally, our native space is the largest function space over which the proposed regularization functional $f \mapsto \|\mathrm{L_R}\{f\}\|_{\mathcal{X}'}$ is well-defined. The precise description of $\mathcal{X}'_{\mathrm{L_R}}(\mathbb{R}^d)$, however, is a bit more involved. As first step, we need to restrict the dual pair $(\mathcal{X}, \mathcal{X}')$ to the range of the (filtered) Radon transform. This yields the Banach spaces $(\mathcal{X}_{\mathrm{Rad}}, \mathcal{X}'_{\mathrm{Rad}})$, as defined by (36) and (37), with the pairs of interest being $(C_{0,\mathrm{Rad}}, \mathcal{M}_{0,\mathrm{Rad}})$ and $(L_{q,\mathrm{Rad}}, L_{p,\mathrm{Rad}})$ with $\frac{1}{p} + \frac{1}{q} = 1$ and $p \in [1, \infty)$. Given some spline-admissible operator L (Definition 2 and 3), we then define our regularization operator and its adjoint as

$$\mathrm{L_R} \stackrel{\Delta}{=} \mathrm{K}_{\mathrm{rad}}\mathrm{RL} : \mathcal{X}'_{\mathrm{L_R}}(\mathbb{R}^d) \to \mathcal{X}'_{\mathrm{Rad}},$$
$$\mathrm{L}^*_{\mathrm{R}} = \mathrm{L}^*\mathrm{R}^*\mathrm{K}_{\mathrm{rad}} : \mathcal{X}_{\mathrm{Rad}} \to \mathcal{X}_{\mathrm{L_R}}(\mathbb{R}^d)$$

where R denotes the Radon transform and $\mathrm{K}_{\mathrm{rad}}$ is the (self-adjoint) filtering operator such that $\mathrm{K}_{\mathrm{rad}}\mathrm{RR}^* = \mathrm{Id}$ on $\mathcal{X}'_{\mathrm{Rad}}$ (Theorem 3). In order to establish isometries, one needs to be able to invert $\mathrm{L}: \mathcal{X}'_{\mathrm{L_R}}(\mathbb{R}^d) \to \mathrm{R}^*(\mathcal{X}'_{\mathrm{Rad}})$ as well as $\mathrm{L_R}$, which is feasible if one factors out the null space $\mathcal{P}$, which is common to both. This motivates us to define the directed inverse operators

$$\mathrm{L}_{\mathcal{P}}^{-1} \triangleq (\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}})\mathrm{L}^{-1} : \mathrm{R}^*\big(\mathcal{X}'_{\mathrm{Rad}}\big) \to \mathcal{X}'_{\mathrm{L_R}}(\mathbb{R}^d) \qquad \text{(Right inverse of L)}$$
$$\mathrm{L}_{\mathcal{P}}^{-1*} = \mathrm{L}^{-1*}(\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}'}) : \mathcal{X}_{\mathrm{L_R}}(\mathbb{R}^d) \to \mathrm{K_{rad}}\mathrm{R}(\mathcal{X}_{\mathrm{Rad}}) \quad \text{(Left inverse of L}^*\text{)}$$
$$\mathrm{L}_{\mathrm{R}}^{\dagger} \triangleq \mathrm{L}_{\mathcal{P}}^{-1}\mathrm{R}^* : \mathcal{X}'_{\mathrm{Rad}} \to \mathcal{X}'_{\mathrm{L_R}}(\mathbb{R}^d) \tag{47}$$
$$\mathrm{L}_{\mathrm{R}}^{*\dagger} = \mathrm{R}\mathrm{L}_{\mathcal{P}}^{-1*} : \mathcal{X}_{\mathrm{L_R}}(\mathbb{R}^d) \to \mathcal{X}_{\mathrm{Rad}} \tag{48}$$

where the operators $\mathrm{L}_{\mathrm{R}}^{\dagger}$ and $\mathrm{L}_{\mathrm{R}}^{*\dagger}$ are generalized inverses[3] of $\mathrm{L_R}$ and $\mathrm{L}_{\mathrm{R}}^*$, respectively. We now have all the ingredients to specify our native space as

$$\begin{aligned}\mathcal{X}'_{\mathrm{L_R}}(\mathbb{R}^d) &= \mathrm{L}_{\mathrm{R}}^{\dagger}\big(\mathcal{X}'_{\mathrm{Rad}}\big) \oplus \mathcal{P}\\ &= \{f \in L_{\infty,-n_0}(\mathbb{R}^d) : \|\mathrm{L_R}\{f\}\|_{\mathcal{X}'} + \|\mathrm{Proj}_{\mathcal{P}}\{f\}\|_{\mathcal{P}} < \infty\}\\ &= \{\mathrm{L}_{\mathrm{R}}^{\dagger}\{w\} + p_0 : \quad (w, p_0) \in \mathcal{X}'_{\mathrm{Rad}} \times \mathcal{P}\},\end{aligned} \tag{49}$$

which is isometrically isomorphic to $\mathcal{X}'_{\mathrm{Rad}} \times \mathcal{P}$, as expressed by (49). The key property there is that $\mathrm{L_R}\mathrm{L}_{\mathrm{R}}^{\dagger} = \mathrm{Id}$ on $\mathcal{X}'_{\mathrm{Rad}}$, while $\mathrm{L_R}\{p_0\} = 0$ for all $p_0 \in \mathcal{P}$ (Theorem 6). Moreover, $\mathcal{X}'_{\mathrm{L_R}}(\mathbb{R}^d)$ is the topological dual of the *predual space*

$$\begin{aligned}\mathcal{X}_{\mathrm{L_R}}(\mathbb{R}^d) &= \mathrm{L}_{\mathrm{R}}^*\big(\mathcal{X}_{\mathrm{Rad}}\big) \oplus \mathcal{P}'\\ &= \{\nu \in \mathcal{S}'(\mathbb{R}^d) : \|\nu\|_{\mathcal{X}_{\mathrm{L_R}}} = \max(\|\mathrm{L}_{\mathrm{R}}^{*\dagger}\{\nu\}\|_{\mathcal{X}}, \|\mathrm{Proj}_{\mathcal{P}'}\{\nu\}\|_{\mathcal{P}'}) < \infty\}\\ &= \{\mathrm{L}_{\mathrm{R}}^*\{v\} + p_0^* : \quad (v, p_0^*) \in \mathcal{X}_{\mathrm{Rad}} \times \mathcal{P}'\},\end{aligned} \tag{50}$$

which is a Banach space, as shown in Theorem 9. The validity of this dual pairing can be checked formally in the absence of null space components: For any $(f, \nu) \in \mathcal{X}'_{\mathrm{L_R}}(\mathbb{R}^d) \times \mathcal{X}_{\mathrm{L_R}}(\mathbb{R}^d)$ with $\mathrm{Proj}_{\mathcal{P}}\{f\} = 0$ and $\mathrm{Proj}_{\mathcal{P}'}\{\nu\} = 0$, we have that

$$\langle f, \nu\rangle = \langle \mathrm{L}_{\mathrm{R}}^{\dagger}\{w\}, \mathrm{L}_{\mathrm{R}}^*\{v\}\rangle = \langle \mathrm{L_R}\mathrm{L}_{\mathrm{R}}^{\dagger}\{w\}, v\rangle_{\mathrm{Rad}} = \langle w, v\rangle_{\mathrm{Rad}}$$

with $(w, v) \in \mathcal{X}'_{\mathrm{Rad}} \times \mathcal{X}_{\mathrm{Rad}}$. Finally, since $\mathrm{Proj}_{\mathcal{P}}$ continuously maps $\mathcal{X}'_{\mathrm{L_R}}(\mathbb{R}^d) \to \mathcal{P} \xhookrightarrow{\text{iso.}} \mathcal{X}'_{\mathrm{L_R}}(\mathbb{R}^d)$, we can identify $\mathrm{Proj}_{\mathcal{P}'}$ as its adjoint $(\mathrm{Proj}_{\mathcal{P}})^* = \mathrm{Proj}_{\mathcal{P}'}$: $\mathcal{X}_{\mathrm{L_R}}(\mathbb{R}^d) \to \mathcal{P}' \xhookrightarrow{\text{iso.}} \mathcal{X}_{\mathrm{L_R}}(\mathbb{R}^d)$.

To ensure that the generic regression problem in Theorem 5 is well-defined, we also need a mild hypothesis on the structure of the data points.

Definition 7 (Admissible data points) Let $N_0 = \dim\mathcal{P}$ where $\mathcal{P} = \mathcal{P}_{n_0}$ is the polynomial null space of $\mathrm{L_R}$. Then, the set of data points $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_M\} \subset \mathbb{R}^d$ with $M \geq N_0$ is said to be $\mathcal{P}$-*admissible* if the sampling matrix $\mathbf{H} = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_M\end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^{M \times N_0}$ with $\mathbf{v}_m = \big(\langle \delta(\cdot - \boldsymbol{x}_m), m_{\boldsymbol{k}}\rangle\big)_{|\boldsymbol{k}| \leq n_0} \in \mathbb{R}^{N_0}$ is of rank N_0.

[3] The precise properties of these inverse operators are stated in Theorem 6 and 9 .

The rank condition is precisely the condition under which the classical least-squares polynomial fitting problem

$$\min_{f\in\mathcal{P}} \sum_{m=1}^{M} |y_m - f(\boldsymbol{x}_m)|^2 = \min_{\mathbf{b}\in\mathbb{R}^{N_0}} \|\mathbf{y} - \mathbf{H}\mathbf{b}\|^2 \tag{51}$$

is well-posed. Indeed, the formal solution of (51) is $p_0(\boldsymbol{x}) = \sum_{|\boldsymbol{k}|\le n_0} b_{\boldsymbol{k}} m_{\boldsymbol{k}}(\boldsymbol{x})$ with $\mathbf{b} = (b_{\boldsymbol{k}}) = (\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}\mathbf{y}$ where the rank condition guarantees the invertibility of the normal matrix $(\mathbf{H}^\mathsf{T}\mathbf{H}) \in \mathbb{R}^{N_0\times N_0}$.

We are now ready to formulate our extended representer theorem, which supports a rich panorama of regression models. A case of direct practical relevance is obtained by setting $\mathrm{L} = (-\Delta)^{\frac{\alpha+1}{2}}$ (fractional Laplacian) with $\widehat{L}_{\mathrm{rad}}(\omega) = |\omega|^{\alpha+1}$. Indeed, we shall see that, for $\mathcal{X} = L_2$, this essentially yield the classical thin-plate splines (see Section 4.2), while, for $\mathcal{X}' = \mathcal{M}$, it results in neural networks with activation functions (ρ_{rad}) labelled as "ridge splines" (including ReLU) and "fractional splines" in Table 1 (Section 4.5).

Theorem 5 *Let us consider the following setting.*

- *A proper, lower-semicontinuous, coercive, and convex loss functional* $E : \mathbb{R}\times\mathbb{R} \to \mathbb{R}^+ \cup \{+\infty\}$.
- *An isotropic, spline-admissible operator* L *with frequency profile* $\widehat{L}_{\mathrm{rad}}$ *and polynomial null space* $\mathcal{P}_{n_0}$ *of degree* n_0, *possibly trivial with the convention that* $\mathcal{P}_{-1} = \{0\}$.
- *A convex, increasing function* $\psi : \mathbb{R}^+ \to \mathbb{R}^+$.
- *A set* $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_M\} \subset \mathbb{R}^d$ *of* $\mathcal{P}_{n_0}$*-admissible data points.*

Then, for any fixed $\mathbf{y} = (y_m) \in \mathbb{R}^M$, *the generic functional-optimization problem*

$$S = \arg \min_{f\in\mathcal{X}'_{\mathrm{L_R}}(\mathbb{R}^d)} \left(\sum_{m=1}^{M} E(y_m, f(\boldsymbol{x}_m)) + \psi(\|\mathrm{L_R}\{f\}\|_{\mathcal{X}'}) \right), \tag{52}$$

with $\mathrm{L_R} = \mathrm{K_{rad}}\mathrm{R}\mathrm{L}$, *and* ψ, $\mathcal{X}'$ *as stated below, always has a solution.*

1. *When* $\mathcal{X} = \mathcal{X}' = L_2(\mathbb{R}\times\mathbb{S}^{d-1})$ *and* ψ *is strictly convex, the solution of* (52) *is unique and representable by the linear kernel expansion*

$$f(\boldsymbol{x}) = p_0(\boldsymbol{x}) + \sum_{m=1}^{M} a_m \rho_{\mathrm{iso}}(\boldsymbol{x} - \boldsymbol{x}_m), \tag{53}$$

where $\rho_{\mathrm{iso}} = 2(2\pi)^{d-1}\mathcal{F}^{-1}\{1/(|\widehat{L}_{\mathrm{rad}}(\|\boldsymbol{\omega}\|)|^2\|\boldsymbol{\omega}\|^{d-1})\}$ *is a radial basis function,* $(a_m) \in \mathbb{R}^M$ *an adequate set of coefficients, and* $p_0 \in \mathcal{P}_{n_0}$ *a polynomial that lies in the null space of* $\mathrm{L_R}$.

2. *When* $\mathcal{X}' = L_p(\mathbb{R} \times \mathbb{S}^{d-1})$ *with* $p \in (1,2]$ *and* ψ *is strictly convex, the solution is unique and admits the parametric representation*

$$f(\boldsymbol{x}) = p_0(\boldsymbol{x}) + \mathrm{L}_{\mathrm{R}}^{\dagger} \circ \mathrm{J}_q\{\sum_{m=1}^{M} a_m \nu_{\boldsymbol{x}_m}\}(\boldsymbol{x}) \tag{54}$$

with basis functions $\nu_{\boldsymbol{x}_1}, \dots, \nu_{\boldsymbol{x}_M} \in L_q(\mathbb{R} \times \mathbb{S}^{d-1})$ *specified by* (70) *and parameters* $(a_m) \in \mathbb{R}^M$, $p_0 \in \mathcal{P}_{n_0}$, *and where* J_q *is the pointwise nonlinearity given by* (92) *with* $q = p/(p-1)$. *(The latter is the duality map* $\mathrm{J}_{\mathcal{X}} : \mathcal{X} \to \mathcal{X}'$ *with* $\mathcal{X} = L_q(\mathbb{R} \times \mathbb{S}^{d-1})$*—see Appendix A for explanations).*

3. *When* $\mathcal{X}' = \mathcal{M}(\mathbb{R} \times \mathbb{S}^{d-1})$ *and* ψ *is strictly increasing, the solution set is the weak∗-closed convex hull of its extreme points, which are all of the form*

$$f_{\mathrm{ext}}(\boldsymbol{x}) = p_0(\boldsymbol{x}) + \sum_{k=1}^{K_0} a_k \rho_{\mathrm{rad}}(\boldsymbol{\xi}_k^{\mathsf{T}} \boldsymbol{x} - \tau_k) \tag{55}$$

with activation function $\rho_{\mathrm{rad}} = \mathcal{F}^{-1}\{1/\widehat{L}_{\mathrm{rad}}\}$, *for some* $K_0 \leq M - \dim \mathcal{P}_{n_0}$, $(a_k, \boldsymbol{\xi}_k, \tau_k) \in \mathbb{R} \times \mathbb{S}^{d-1} \times \mathbb{R}$ *for* $k = 1, \dots, K_0$, *and a null-space component* $p_0 \in \mathcal{P}_{n_0}$. *The optimal regularization cost associated with* (55) *is* $\|\mathrm{L}_{\mathrm{R}} f_{\mathrm{ext}}\|_{\mathcal{M}} = \sum_{k=1}^{K_0} |a_k|$ *and is shared by all solutions.*

Proof Since $\mathrm{L}_{\mathrm{R}}^*$ is injective on $\mathcal{S}_{\mathrm{Rad}}$ and, by extension, on the completed space $\mathcal{X}_{\mathrm{Rad}}$, the image space $\mathcal{U} = \mathrm{L}_{\mathrm{R}}^*(\mathcal{X}_{\mathrm{Rad}})$ is complete as well (see proof of Theorem 9 for the details of the construction of $\mathcal{U}$). Its continuous dual is given by $\mathcal{U}' = \mathrm{L}_{\mathcal{P}}^{-1}\mathrm{R}^*(\mathcal{X}'_{\mathrm{Rad}})$, in reason of the identities $\mathrm{R}\mathrm{L}_{\mathcal{P}}^{-1*}\mathrm{L}_{\mathrm{R}}^* = \mathrm{R}\mathrm{L}_{\mathcal{P}}^{-1*}\mathrm{L}^*\mathrm{R}^*\mathrm{K}_{\mathrm{rad}} = \mathrm{R}\mathrm{R}^*\mathrm{K}_{\mathrm{rad}} = \mathrm{Id}$ on $\mathcal{X}_{\mathrm{Rad}}$. Likewise, $\mathcal{P}'$, as identified by (15), is a finite-dimensional Banach space. Its dual is simply $(\mathcal{P}')' = \mathcal{P}$ (the null space of both L and L_{R}), owing to the property that all finite-dimensional spaces are reflexive. Using the notation for direct-sum topologies of [58], we then observe that $\mathcal{X}_{\mathrm{L}_{\mathrm{R}}}(\mathbb{R}^d) = (\mathcal{U} \oplus \mathcal{P}')_{\ell_\infty}$, whose formal dual $(\mathcal{U} \oplus \mathcal{P}')'_{\ell_\infty} = (\mathcal{U}' \oplus \mathcal{P})_{\ell_1}$ is precisely the native space $\mathcal{X}'_{\mathrm{L}_{\mathrm{R}}}(\mathbb{R}^d)$ described by (49). By writing $f = \mathrm{L}_{\mathcal{P}}^{-1}\mathrm{R}^*\{w\} + p_0$ and recalling that $\mathrm{L}\mathrm{L}_{\mathcal{P}}^{-1} = \mathrm{Id}$ (right-inverse property), we then identify the regularization functional as

$$\begin{aligned}
\|\mathrm{L}_{\mathrm{R}} f\|_{\mathcal{X}'} &= \|\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{\mathrm{L}\mathrm{L}_{\mathcal{P}}^{-1}\mathrm{R}^* w + \mathrm{L}p_0\}\|_{\mathcal{X}'} \\
&= \|\mathrm{K}_{\mathrm{rad}}\mathrm{R}\mathrm{R}^*\{w\} + \mathrm{K}_{\mathrm{rad}}\mathrm{R}\{0\}\|_{\mathcal{X}'} \\
&= \|w\|_{\mathcal{X}'_{\mathrm{Rad}}} = \|\mathrm{Proj}_{\mathcal{U}'} f\|_{\mathcal{U}'}
\end{aligned}$$

which, in effect, converts the seminorm over $\mathcal{X}'_{\mathrm{L}_{\mathrm{R}}}(\mathbb{R}^d)$ into a norm over $\mathcal{U}'$ by factoring out the null space of L_{R}. The other mathematical ingredient for the optimization problem (52) to be well-posed is the weak∗ continuity of the sampling functionals $\delta(\cdot - \boldsymbol{x}_m) : f \mapsto f(\boldsymbol{x}_m)$ in the underlying topology. This is equivalent to $\delta(\cdot - \boldsymbol{x}_m) \in \mathcal{X}_{\mathrm{L}_{\mathrm{R}}}(\mathbb{R}^d)$ for any $\boldsymbol{x}_m \in \mathbb{R}^d$. In the present context, this condition can be reframed as $\nu_{\boldsymbol{x}_m} = \mathrm{L}_{\mathrm{R}}^{*\dagger}\{\delta(\cdot - \boldsymbol{x}_m)\} \in \mathcal{X}_{\mathrm{Rad}}$, which is a property that is established in Theorem 7 for the cases $\mathcal{X} = C_0$ and $\mathcal{X} = L_q$ for any $q \geq 2$.

The existence of the solution and the parametric descriptions (53), (54), and (55) then follows from the three cases of the abstract representer theorem for direct sums [58, Theorem 3]. The link with the abstract theorem is made by identifying $\mathcal{N}_{\boldsymbol{p}} = \mathcal{P}$, $\mathcal{U} \oplus \mathcal{N}_{\boldsymbol{p}^*} = \mathcal{X}_{\mathrm{L_R}}(\mathbb{R}^d)$, $\mathcal{U}' \oplus \mathcal{N}_{\boldsymbol{p}} = \mathcal{X}'_{\mathrm{L_R}}(\mathbb{R}^d)$ and $\nu_m = \delta(\cdot - \boldsymbol{x}_m)$ for $m = 1, \dots, M$. As for the required technical assumptions, they directly follow from the weak$*$ continuity of the sampling functionals (i.e., $\nu_m \in \mathcal{U} \oplus \mathcal{N}_{\boldsymbol{p}^*}$) and the $\mathcal{P}$-admissibility hypothesis in Definition 7, with the $\mathbf{v}_m$ being the same as in the statement of the abstract theorem. The duality map that is required for the first and second scenarios is $\mathrm{J}_{\mathcal{U}} = \mathrm{L}_{\mathrm{R}}^{\dagger} \circ \mathrm{J} \circ \mathrm{L}_{\mathrm{R}}^{*\dagger} : \mathcal{U} \to \mathcal{U}'$ (see Appendix A, Proposition 4 with $\mathrm{T} = \mathrm{L}_{\mathrm{R}}^*$, $\mathrm{T}^{-1} = \mathrm{L}_{\mathrm{R}}^{*\dagger} = \mathrm{R}^*\mathrm{L}_{\mathcal{P}}^{*-1}$, and $(\mathrm{T}^*)^{-1} = \mathrm{L}_{\mathrm{R}}^{\dagger} = \mathrm{L}_{\mathcal{P}}^{-1}\mathrm{R}^*$).

To describe the solution set for the non-reflexive case $\mathcal{X}' = \mathcal{M}$, we invoke the third case of [58, Theorem 3], which tells us that the extreme points of S can all be expressed as the sum of a null-space component plus a linear combination of $K_0 \leq M - \dim(\mathcal{P})$ atoms e_k that are selected adaptively within a dictionary consisting of the extreme points of the unit ball in $\mathcal{U}'$: $B(\mathcal{U}') = \{u \in \mathcal{U}' : \|u\|_{\mathcal{U}'} \leq 1\}$. Because of the linear isomorphism $\mathcal{U}' = \mathrm{L}_{\mathrm{R}}^{\dagger}\big(\mathcal{M}_{\mathrm{Rad}}\big)$, $\mathrm{ext}B(\mathcal{U}) = \mathrm{L}_{\mathrm{R}}^{\dagger}\big(\mathrm{ext}B(\mathcal{M}_{\mathrm{Rad}})\big)$. Next, we use the property that $\mathcal{M}_{\mathrm{Rad}} = \mathcal{M}_{\mathrm{even}}(\mathbb{R} \times \mathbb{S}^{d-1})$ whose extreme points are $\{\pm\delta_{\mathrm{Rad},\boldsymbol{z}}\}_{\boldsymbol{z} \in \mathbb{P}^d}$. Each $\delta_{\mathrm{Rad},\boldsymbol{z}_k} \in \mathrm{ext}B(\mathcal{M}_{\mathrm{Rad}})$ then bijectively maps into an extreme point $e_k = \mathrm{L}_{\mathrm{R}}^{\dagger}\{\delta_{\mathrm{Rad},\boldsymbol{z}_k}\} \in \mathrm{ext}B(\mathcal{U}')$, and vice versa. Finally, by recalling that $\mathrm{L}_{\mathrm{R}}^{\dagger} = (\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}})\mathrm{L}^{-1}\mathrm{R}^*$ and by invoking Theorem 4 to show that $\mathrm{L}^{-1}\mathrm{R}^*\{\delta_{\mathrm{Rad},(t_k,\boldsymbol{\xi}_k)}\} = \mathrm{L}^{-1}\mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{\delta(\boldsymbol{\xi}_k^{\mathsf{T}}\boldsymbol{x} - t_k)\} = \mathrm{L}^{-1}\{\delta(\boldsymbol{\xi}_k^{\mathsf{T}}\boldsymbol{x} - t_k)\} = \rho_{\mathrm{rad}}(\boldsymbol{\xi}_k^{\mathsf{T}}\boldsymbol{x} - t_k)$, we find that

$$e_k = \pm\rho_{\mathrm{rad}}(\boldsymbol{\xi}_k^{\mathsf{T}}\boldsymbol{x} - t_k) \mp p_{0,k}, \tag{56}$$

where $p_{0,k} = \mathrm{Proj}_{\mathcal{P}}\{\rho_{\mathrm{rad}}(\boldsymbol{\xi}_k^{\mathsf{T}}\boldsymbol{x} - \tau_k)\} \in \mathcal{P}$. Since every extreme point of $B(\mathcal{U}')$ is necessarily of the form (56), we can substitute this expression in the generic expansion $f_{\mathrm{ext}} = \tilde{p}_0 + \sum_{k=1}^{K_0} a_k e_k$ which, upon collection of all null-space components, yields (55). □

The two cases in Theorem 5 that are of direct practical relevance to machine learning are Items 1 and 3. The first scenario yields a learning architecture that is a linear expansion of RBFs, which also has a classical RKHS interpretation, as explained in Section 4.2.

The form of the solution in Item 3 is equivalent to a shallow network with the weights $(\boldsymbol{\xi}_k)$ of the hidden layer being normalized. It actually turns out that this normalization is inconsequential when the activation is a homogeneous function. This happens for instance when the regularization operator $\mathrm{L} = (-\Delta)^{\frac{\alpha+1}{2}}$ is a fractional Laplacian which maps into $\rho_{\mathrm{Rad}}(t) \propto |t|^{\alpha}$. Indeed, for any $(\boldsymbol{w}_k, b_k) \in \mathbb{R}^d \times \mathbb{R}$, we have that

$$|\boldsymbol{w}_k^{\mathsf{T}}\boldsymbol{x} - b_k|^{\alpha} = \|\boldsymbol{w}_k\|^{\alpha}|\boldsymbol{\xi}_k^{\mathsf{T}}\boldsymbol{x} - t_k|^{\alpha} \tag{57}$$

with $\boldsymbol{\xi}_k = \boldsymbol{w}_k/\|\boldsymbol{w}_k\|$ and $t_k = b_k/\|\boldsymbol{w}_k\|$, which indicates that the normalization (or lack thereof) can be absorbed in the weights (a_k) of the output layer. The case $\alpha = 1$ with $\mathrm{L} = \Delta$ (Laplacian) is particularly attractive, as it nicely maps into a ReLU network with one hidden layer and a skip connection to implement the affine component [37].

The form of the solution in Item 2 is more involved, but still useful to get insight into the transition with p from RBFs to neural nets. The equivalence with Item 1 for $p = 2$ is demonstrated in the Section 4.2. As for the behavior as $p \to 1$, we observe that the effect of the duality map J_q as $q \to \infty$ is to pull a few maximal values to infinity, while attenuating all other (non-supremum) values towards zero. In effect, this means that $\mathrm{J}_q\{\sum_{m=1}^{M} a_m \nu_{\boldsymbol{x}_m}\}$ will exhibit peaks that become more and more pronounced, and eventually converge to a sum of impulses as $p \to 1$, which is consistent with the limit form given by (55).

4.2 Connection with RKHS Methods

The scenario $\mathcal{X} = L_2$ in Theorem 5 is compatible with the kernel models of "classical" machine learning [40, 50]. This is because the underlying native space is a reproducing-kernel Hilbert space whose topological structure is now made explicit.

Proposition 2 (Characterization of RKHS for $\mathcal{X} = L_2$) *Let* L *be a spline admissible operator with a polynomial null space of degree n_0 and consider the self-adjoint operator* $\mathrm{A} = (\mathrm{L}^*\mathrm{KL})^{-1}$. *Then, the native space $\mathcal{H} = L'_{2,\mathrm{L_R}}(\mathbb{R}^d)$, defined by* (49) *with $\mathcal{X}' = \mathcal{X} = L_2$, is the reproducing-kernel Hilbert space $\mathcal{H} = \mathcal{U}' \oplus \mathcal{P}$ associated with the composite inner product*

$$\begin{aligned}\langle f_1, f_2\rangle_{\mathcal{H}} &= \langle \mathrm{L_R}\{f_1\}, \mathrm{L_R}\{f_2\}\rangle_{\mathrm{Rad}} + \sum_{|\boldsymbol{k}|\le n_0} \langle m^*_{\boldsymbol{k}}, f_1\rangle\langle m^*_{\boldsymbol{k}}, f_2\rangle \\ &= \langle (\mathrm{L}^*\mathrm{KL})\{f_1\}, f_2\rangle + \sum_{|\boldsymbol{k}|\le n_0} \langle m^*_{\boldsymbol{k}}, f_1\rangle\langle m^*_{\boldsymbol{k}}, f_2\rangle\end{aligned} \tag{58}$$

whose leading term can also be written as

$$\begin{aligned}\langle (\mathrm{L}^*\mathrm{KL})\{f_1\}, f_2\rangle &= \langle (\mathrm{L}^*\mathrm{KL})\mathrm{Proj}_{\mathcal{U}'}\{f_1\}, \mathrm{Proj}_{\mathcal{U}'}\{f_2\}\rangle \\ &= \langle \mathrm{Proj}_{\mathcal{U}'}\{f_1\}, \mathrm{Proj}_{\mathcal{U}'}\{f_2\}\rangle_{\mathcal{U}'},\end{aligned} \tag{59}$$

where $\mathrm{Proj}_{\mathcal{U}'} = \mathrm{Id} - \mathrm{Proj}_{\mathcal{P}} : \mathcal{H} \to \mathcal{U}'$. *The topological dual of $\mathcal{H}$ is the Hilbert space $\mathcal{H}' = L_{2,\mathrm{L_R}}(\mathbb{R}^d) = \mathcal{U} \oplus \mathcal{P}'$ equipped with the inner product*

$$\begin{aligned}\langle g_1, g_2\rangle_{\mathcal{H}'} &= \langle \mathrm{L}_{\mathrm{R}}^{*\dagger}\{g_1\}, \mathrm{L}_{\mathrm{R}}^{*\dagger}\{g_2\}\rangle_{\mathrm{Rad}} + \sum_{|\boldsymbol{k}|\le n_0} \langle m_{\boldsymbol{k}}, g_1\rangle\langle m_{\boldsymbol{k}}, g_2\rangle \\ &= \langle \mathrm{L}_{\mathrm{R}}^{\dagger}\mathrm{L}_{\mathrm{R}}^{*\dagger}\{g_1\}, g_2\rangle + \sum_{|\boldsymbol{k}|\le n_0} \langle m_{\boldsymbol{k}}, g_1\rangle\langle m_{\boldsymbol{k}}, g_2\rangle \\ &= \langle \mathrm{A}\mathrm{Proj}_{\mathcal{U}}\{g_1\}, \mathrm{Proj}_{\mathcal{U}}\{g_2\}\rangle + \sum_{|\boldsymbol{k}|\le n_0} \langle m_{\boldsymbol{k}}, g_1\rangle\langle m_{\boldsymbol{k}}, g_2\rangle,\end{aligned} \tag{60}$$

where $\mathrm{Proj}_{\mathcal{U}} = \mathrm{Id} - \mathrm{Proj}_{\mathcal{P}'} : \mathcal{H}' \to \mathcal{U}$. *The corresponding linear isometries (Riesz maps) that map* $\mathcal{U} \to \mathcal{U}'$, $\mathcal{P}' \to \mathcal{P}$, *and* $\mathcal{H} \to \mathcal{H}'$ *are*

$$\begin{aligned}
\mathrm{J}_{\mathcal{U}} &= \mathrm{A} : \mathcal{U} \to \mathcal{U}' \\
&= (\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}})\mathrm{A}(\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}'}) : \mathcal{U} \oplus \mathcal{P}' \to \mathcal{U}', \\
\mathrm{J}_{\mathcal{P}'} &= \mathrm{Proj}_{\mathcal{P}} : \mathcal{P}' \to \mathcal{P}, \\
&= \mathrm{Proj}_{\mathcal{P}}\mathrm{Proj}_{\mathcal{P}'} : \mathcal{U} \oplus \mathcal{P}' \to \mathcal{P}, \\
\mathrm{J}_{\mathcal{H}} &= ((\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}})\mathrm{A}(\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}'}) + \mathrm{Proj}_{\mathcal{P}}\mathrm{Proj}_{\mathcal{P}'}) : \mathcal{H} \to \mathcal{H}',
\end{aligned}$$

where the second, alternative forms of $\mathrm{J}_{\mathcal{U}}$ *and* $\mathrm{J}_{\mathcal{P}'}$ *that include projectors are the proper extension of those operators to the whole space* $\mathcal{H} = \mathcal{U}' \oplus \mathcal{P}$.

Proposition 2 is obtained as a corollary of Theorems 6 and 9 with $\mathcal{X} = \mathcal{X}' = L_2$, with the help of the intertwining property $\mathrm{L_R} = \mathrm{K_{rad}RL} = \mathrm{RKL}$ (see the discussion at the end of Section 3.2). The technical part is to establish the completeness of $\mathcal{U}$ (resp., $\mathcal{H}' = \mathcal{U} \oplus \mathcal{P}'$), which then implies that of $\mathcal{U}'$ (resp., $\mathcal{H} = \mathcal{H}'' = \mathcal{U}' \oplus \mathcal{P}$) by duality. As for the Hilbert-space property, the maps defined by (58) and (60) are obviously bilinear and symmetric. To show that (58) is also positive definite, we invoke Theorem 6, which states that any $f \in \mathcal{H} = \mathcal{U}' \oplus \mathcal{P}$ has a unique decomposition as $f = \mathrm{L}_{\mathrm{R}}^{\dagger}\{w\} + p_0$, with $w = \mathrm{L_R}\{f\} \in \mathcal{X}'$ and $p_0 = \mathrm{Proj}_{\mathcal{P}}\{f\} = \sum_{|\boldsymbol{k}| \le n_0} b_{\boldsymbol{k}} m_{\boldsymbol{k}} \in \mathcal{P}$ with $b_{\boldsymbol{k}} = \langle m_{\boldsymbol{k}}^*, f\rangle$. This then yields that

$$\langle f, f\rangle_{\mathcal{H}} = \langle \mathrm{L_R}\{f\}, \mathrm{L_R}\{f\}\rangle + \sum_{|\boldsymbol{k}| \le n_0} |b_{\boldsymbol{k}}|^2 = \underbrace{\langle w, w\rangle_{L_2(\mathbb{R}\times\mathbb{S}^{d-1})}}_{\|w\|^2_{\mathcal{X}'}} + \underbrace{\|\boldsymbol{b}\|_2^2}_{\|p_0\|^2_{\mathcal{P}}} \ge 0,$$

which vanishes if and only if $f = 0$. Likewise, one readily verifies that the semi-inner products associated with each individual term in (60) induce the two component norms $\|v\|_{\mathcal{X}}$ and $\|p_0^*\|_{\mathcal{P}'}$ for $\mathcal{X} = L_2$ that appear in the definition (50) of the predual space.

The denomination RKHS applies to any Hilbert space $\mathcal{H}$ of functions on $\mathbb{R}^d$ such that $\delta(\cdot - \boldsymbol{x}_0) \in \mathcal{H}'$ for any $\boldsymbol{x}_0 \in \mathbb{R}^d$. In our case, this property is equivalent to $\mathrm{L}_{\mathrm{R}}^{*\dagger}\{\delta(\cdot - \boldsymbol{x}_0)\} \in L_{2,\mathrm{Rad}}$, which follows from Theorem 7.

To get further insight, we now show that the RBF solution (53) is a particular case of the L_p solution (54) with $p = 2$. Since $L_2 = (L_2)'$ is its own dual, $\mathrm{J} = \mathrm{Id}$, which allows us to manipulate (54) as

$$
\begin{aligned}
f &= p_0 + \mathrm{L}_{\mathcal{P}}^{-1}\mathrm{R}^*\mathrm{R}\mathrm{L}_{\mathcal{P}}^{-1*}\{\sum_{m=1}^{M} a_m \delta(\cdot - \boldsymbol{x}_m)\} \\
&= p_0 + \mathrm{L}_{\mathcal{P}}^{-1}\mathrm{K}^{-1}\mathrm{L}_{\mathcal{P}}^{-1*}\{\sum_{m=1}^{M} a_m \delta(\cdot - \boldsymbol{x}_m)\} \\
&= p_0 + (\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}})\mathrm{A}(\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}'})\{\sum_{m=1}^{M} a_m \delta(\cdot - \boldsymbol{x}_m)\} \qquad (61) \\
&= p_0 + \mathrm{A}\{\sum_{m=1}^{M} a_m \delta(\cdot - \boldsymbol{x}_m)\} = p_0 + \sum_{m=1}^{M} a_m \rho_{\mathrm{iso}}(\boldsymbol{x} - \boldsymbol{x}_m), \qquad (62)
\end{aligned}
$$

where $\rho_{\mathrm{iso}} = (\mathrm{L}^*\mathrm{KL})^{-1}\{\delta\} = \mathrm{A}\{\delta\} : \mathbb{R}^d \to \mathbb{R}$ is the Green's function of $(\mathrm{L}^*\mathrm{KL})$ and $p_0 \in \mathcal{P}$. The nonobvious simplification from (61) to (62) results from two crucial observations: (1) the "orthogonality-to-the-null-space" condition $\sum_{m=1}^{M} a_m \delta(\cdot - \boldsymbol{x}_m) \in \mathcal{U}$ is necessary for optimality; and (2) the availability of the identity

$$
\forall u \in \mathcal{U}: \quad (\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}})\mathrm{A}(\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}'})\{u\} = \mathrm{A}\{u\} = u^* \in \mathcal{U}',
$$

which is tightly linked to the specification of the underlying spaces in Proposition 2. Likewise, we find that the quadratic regularization cost associated with the linear model (62) is $\boldsymbol{a}^{\mathsf{T}}\mathbf{G}\boldsymbol{a}$, where $\mathbf{G} \in \mathbb{R}^{M\times M}$ is a symmetric, conditionally positive-definite matrix (see [31]) whose entries are calculated as follows:

$$
\begin{aligned}
[\mathbf{G}]_{m,n} &= \langle \mathrm{L}_{\mathrm{R}}\rho_{\mathrm{iso}}(\cdot - \boldsymbol{x}_m), \mathrm{L}_{\mathrm{R}}\rho_{\mathrm{iso}}(\cdot - \boldsymbol{x}_n)\rangle_{\mathrm{Rad}} \\
&= \langle \mathrm{RKL}\{\rho_{\mathrm{iso}}(\cdot - \boldsymbol{x}_m)\}, \mathrm{RKL}\{\rho_{\mathrm{iso}}(\cdot - \boldsymbol{x}_n)\}\rangle_{\mathrm{Rad}} \\
&= \langle \mathrm{KL}\{\rho_{\mathrm{iso}}(\cdot - \boldsymbol{x}_m)\}, \mathrm{R}^*\mathrm{RKL}\{\rho_{\mathrm{iso}}(\cdot - \boldsymbol{x}_n)\}\rangle \\
&= \langle \rho_{\mathrm{iso}}(\cdot - \boldsymbol{x}_m), \mathrm{L}^*\mathrm{K}\underbrace{\mathrm{R}^*\mathrm{RK}}_{\mathrm{Id}}\mathrm{L}\{\rho_{\mathrm{iso}}(\cdot - \boldsymbol{x}_n)\}\rangle \\
&= \langle \rho_{\mathrm{iso}}(\cdot - \boldsymbol{x}_m), \delta(\cdot - \boldsymbol{x}_n)\rangle = \rho_{\mathrm{iso}}(\boldsymbol{x}_n - \boldsymbol{x}_m) = \rho_{\mathrm{iso}}(\boldsymbol{x}_m - \boldsymbol{x}_n).
\end{aligned}
$$

As variant of the L_2 result in Theorem 5, we may also consider the modified regularization operator $\tilde{\mathrm{L}} = \mathrm{K}^{-\frac{1}{2}}\mathrm{L}$ whose frequency response is $\widehat{\tilde{L}}(\boldsymbol{\omega}) = \sqrt{2}(2\pi)^{(d-1)/2}\widehat{L}(\boldsymbol{\omega})/\|\boldsymbol{\omega}\|^{(d-1)/2}$. For this particular setting, $\tilde{\mathrm{L}}_{\mathrm{R}} = \mathrm{RK}^{\frac{1}{2}}\mathrm{L}$, which translates into

$$
\|\tilde{\mathrm{L}}_{\mathrm{R}}\{f\}\|^2_{L_2(\mathbb{R}\times\mathbb{S}^{d-1})} = \|\mathrm{RK}^{\frac{1}{2}}\mathrm{L}\{f\}\|^2_{L_2(\mathbb{R}\times\mathbb{S}^{d-1})} = \|\mathrm{L}\{f\}\|^2_{L_2(\mathbb{R}^d)},
$$

owing to the property that $\mathrm{RK}^{\frac{1}{2}}$ is an L_2 isometry [27]. The proposed Radon-domain regularization therefore reduces to the standard energy functional associated with (semi-)reproducing-kernel Hilbert spaces. The resulting basis function is $\tilde{\rho}_{\mathrm{iso}}(\boldsymbol{x}) = \mathcal{F}^{-1}\{1/|\widehat{L}|^2\}(\boldsymbol{x})$, which is the same as the one encountered in the classical formulation that does not involve the Radon transform.

While this may suggest that the two formulations are equivalent, there is an important difference that concerns the dimension of the null space.

As a matter of illustration, we now compare two schemes that utilize the same "linear" radial basis $\rho_{\mathrm{iso}}(\boldsymbol{x}) \propto \|\boldsymbol{x}\|$. Based on (6), we deduce that this corresponds to the choice $|\widehat{L}(\boldsymbol{\omega})|^2 = \|\boldsymbol{\omega}\|^{d+1}$ in the classical formulation, which induces the regularization norm $\|(-\Delta)^{(d+1)/2}\{f\}\|_{L_2}$ with a polynomial null space of degree $n_0 = \lceil(d+1)/2\rceil > d/2$ [17,63]. In our proposed Radon-domain variant, the appropriate regularization norm is $\|\mathrm{RK}(-\Delta)^{1/2}\{f\}\|_{L_2}$ with a null space of polynomial degree $n_0 = 0$. This solution is attractive because it does not depend on the dimensionality of the data.

4.3 Universal Approximation Properties

The universal-approximation properties of the supervised-learning scheme specified by (52) are supported by Theorem 6 below, which summarizes the properties of the native space $\mathcal{X}'_{\mathrm{L_R}}$ in relation to the regularization operator $\mathrm{L_R}$ and its generalized inverse $\mathrm{L_R^\dagger}$. This result is a direct corollary (dual counterpart) of Theorem 9, whose proof is given in Section 5.2.

Theorem 6 (Properties of the native space $\mathcal{X}'_{\mathrm{L_R}}$) *Let* L *be an isotropic LSI operator that is spline-admissible with a polynomial null space* $\mathcal{P}$ *(possibly trivial) of degree* n_0*. Then, the operators* $\mathrm{L_R} = \mathrm{K_{rad}RL} : \mathcal{X}'_{\mathrm{L_R}} \to \mathcal{X}'_{\mathrm{Rad}}$ *and* $\mathrm{L_R^\dagger} = (\mathrm{Id}-\mathrm{Proj}_{\mathcal{P}})\mathrm{L}^{-1}\mathrm{R}^* : \mathcal{X}'_{\mathrm{Rad}} \to L_{\infty,-n_0}(\mathbb{R}^d)$ *(the adjoint of* $\mathrm{L_R^{*\dagger}}$ *in Theorem 9) are continuous and have the properties*

$$\forall w \in \mathcal{X}'_{\mathrm{Rad}}: \quad \mathrm{L_R L_R^\dagger}\{w\} = w \tag{63}$$

$$\forall p_0 \in \mathcal{P}: \quad \mathrm{L_R}\{p_0\} = 0 \tag{64}$$

$$\forall f \in \mathcal{X}'_{\mathrm{L_R}}(\mathbb{R}^d): \quad \mathrm{L_R^\dagger L_R}\{f\} = (\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}})\{f\} = \mathrm{Proj}_{\mathcal{U}'}\{f\}, \tag{65}$$

where $\mathcal{X}'_{\mathrm{L_R}}(\mathbb{R}^d) \triangleq \mathrm{L_R^\dagger}(\mathcal{X}'_{\mathrm{Rad}}) \oplus \mathcal{P}$ *is equipped with the composite norm* $\|f\|_{\mathcal{X}'_{\mathrm{L_R}}} = \|\mathrm{L_R}\{f\}\|_{\mathcal{X}'_{\mathrm{Rad}}} + \|\mathrm{Proj}_{\mathcal{P}}\{f\}\|_{\mathcal{P}}$*. The space* $\mathcal{X}'_{\mathrm{L_R}}$ *is complete and isomorphic to* $\mathcal{X}'_{\mathrm{Rad}} \times \mathcal{P}$ *with* $f = \mathrm{L_R^\dagger}\{w\} + p_0 \mapsto (w, p_0) = (\mathrm{L_R}\{f\}, \mathrm{Proj}_{\mathcal{P}}\{f\})$*. Moreover,* $\mathcal{S}(\mathbb{R}^d) \hookrightarrow \mathcal{X}'_{\mathrm{L_R}}(\mathbb{R}^d) \hookrightarrow L_{\infty,-n_0}(\mathbb{R}^d) \overset{\mathrm{d.}}{\hookrightarrow} \mathcal{S}'(\mathbb{R}^d)$*.*

To explain how Theorem 6 relates to universality, let us first consider the case $\mathcal{X}' = L_2$ for which we have just shown that $\mathcal{X}'_{\mathrm{L_R}} = \mathcal{H}$ is a RKHS whose (semi-)reproducing kernel is $\rho_{\mathrm{iso}} = (\mathrm{L^*KL})^{-1}\{\delta\}$. From the general properties of RKHS [2,63], we know that $\mathcal{H}$ (as a set) can be specified as $\mathcal{H} = \overline{\mathrm{span}\{\rho_{\mathrm{iso}}(\cdot - \boldsymbol{y})\}_{\boldsymbol{y}\in\mathbb{R}^d}} + \mathcal{P}$, which tells us that the class of RBF estimators of the form given by (53) is dense in $\mathcal{H}$. This means that such estimators can yield an approximation of any $f \in \mathcal{H}$ to an arbitrary degree of precision. In particular, this applies to any $f \in \mathcal{S}(\mathbb{R}^d)$, due to the inclusion $\mathcal{S}(\mathbb{R}^d) \subset \mathcal{X}'_{\mathrm{L_R}}(\mathbb{R}^d)$, as guaranteed by Theorem 6. Now, the key to universal approximation is that $\mathcal{S}(\mathbb{R}^d)$ is dense in most of the classical function spaces

[51], in particular, $C_0(\mathbb{R}^d)$. We then readily deduce that (53), for M sufficiently large and a suitable choice of the $\boldsymbol{x}_m$, has the ability to reproduce any continuous function $f : \mathbb{R}^d \to \mathbb{R}$. This deduction, of course, is consistent with the classical theory of kernel estimators: Our admissibility conditions for $\widehat{L}_{\mathrm{rad}}$ in Definition 2 (resp., in Definition 3) ensure that the function $\rho_{\mathrm{iso}} : \mathbb{R}^d \to \mathbb{R}$ is strictly positive-definite (resp., strictly conditionally positive-definite), which is the standard criterion for universality [31,32,63].

The same kind of density argument can also be made for $\mathcal{X}' = \mathcal{M}$. The relevant basis elements there are the atomic Radon-compatible Dirac measures $\delta_{\mathrm{Rad},(t_k,\boldsymbol{\xi}_k)} \in \mathcal{M}_{\mathrm{Rad}}$ with $(t_k,\boldsymbol{\xi}_k) \in \mathbb{R} \times \mathbb{S}^{d-1}$. These get mapped into $e_k = \mathrm{L}_{\mathrm{R}}^{\dagger}\{\delta_{\mathrm{Rad},(t_k,\boldsymbol{\xi}_k)}\} \in \mathcal{U}'$, which are essentially ridges (up to some polynomial) characterized by $e_k = \rho_{\mathrm{rad}}(\boldsymbol{\xi}_k^{\mathsf{T}} \cdot -t_k) - p_{0,k}$ with $p_{0,k} = \mathrm{Proj}_{\mathcal{P}}\{\rho_{\mathrm{rad}}(\boldsymbol{\xi}_k^{\mathsf{T}} \cdot -t_k)\}$. Thus, by setting $w \approx \sum_k w_k \delta_{\mathrm{Rad},(t_k,\boldsymbol{\xi}_k)}$, we can interpret the generative formula $f = \mathrm{L}_{\mathrm{R}}^{\dagger}\{w\} + p_0$ in Theorem 6 as a linear superposition of ridges plus a global polynomial trend, which is compatible with the form of the estimator in (55). We then invoke the property that $\mathcal{S}(\mathbb{R}^d) \subset \mathcal{M}_{\mathrm{L_R}}(\mathbb{R}^d)$, which implies that any continuous function can be approximated as closely as desired by a member of $\mathcal{M}_{\mathrm{L_R}}(\mathbb{R}^d)$. As it turns out, the latter is representable by a superposition of ridges plus a polynomial of degree n_0. We emphasize that the presence of the polynomial term—the guarantor of stability for Theorem 8—is essential to counterbalance the growth of the individual atoms at infinity. This is an important aspect where our analysis and conclusions deviate from those of [54].

4.4 Regularization Operators for Antisymmetric Activations

While the framework that has been discussed so far is very powerful, it has one shortcoming: it yields canonical activation functions ρ_{rad} that are necessarily symmetric. In some cases such as $\mathrm{L_R} = \mathrm{K}_{\mathrm{rad}}\mathrm{R}(-\Delta)$, these can be converted into one-sided functions such as the ReLU by doctoring the polynomial term. That said, the original scheme that is described by Theorem 5 does not allow for sigmoids, which are frequently used for classification [8]. This is the reason why we now introduce a variant that systematically produces anti-symmetric activations, including sigmoids for $n_0 = 0$.

The idea is to substitute the (symmetric) filtering operator $\mathrm{K}_{\mathrm{rad}}$ by its anti-symmetric counterpart $\tilde{\mathrm{K}}_{\mathrm{rad}}$, which includes an additional Hilbert transform. Specifically, $\tilde{\mathrm{K}}_{\mathrm{rad}}$ is the hyper-spherical radial filter whose frequency response is $\widehat{\tilde{K}}_{\mathrm{rad}}(\omega) = -\mathrm{i}\,\mathrm{sign}(\omega) c_p |\omega|^{-d-1}$ and whose adjoint is $\tilde{\mathrm{K}}_{\mathrm{rad}}^{*} = -\tilde{\mathrm{K}}_{\mathrm{rad}}$. Since the action of the (radial) Hilbert transform $\tilde{\mathrm{H}}_{\mathrm{rad}} : \phi(\cdot,\boldsymbol{\xi}) \mapsto \phi \circledast 1/(\pi t)$ is well-defined on $\mathcal{S}(\mathbb{R} \times \mathbb{S}^{d-1})$ with $\tilde{\mathrm{H}}_{\mathrm{rad}}^{*} = -\tilde{\mathrm{H}}_{\mathrm{rad}} = \tilde{\mathrm{H}}_{\mathrm{rad}}^{-1}$, we have the identity $\mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\mathrm{R} = \mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\tilde{\mathrm{H}}_{\mathrm{rad}}^{*}\tilde{\mathrm{H}}_{\mathrm{rad}}\mathrm{R} = \mathrm{R}^*\tilde{\mathrm{K}}_{\mathrm{rad}}^{*}\tilde{\mathrm{R}} = \mathrm{Id}$ with $\tilde{\mathrm{R}} \triangleq \tilde{\mathrm{H}}_{\mathrm{rad}}\mathrm{R}$. Accordingly, we can essentially replicate the whole mechanism of construction of spaces in Section 3.3 by subtituting $\mathcal{S}_{\mathrm{Rad}}$ by $\tilde{\mathcal{S}}_{\mathrm{Rad}} = \tilde{\mathrm{R}}(\mathcal{S}(\mathbb{R}\times\mathbb{S}^{d-1}) = \tilde{\mathrm{H}}_{\mathrm{rad}}(\mathcal{S}_{\mathrm{Rad}})$, which is a space of odd functions that are smooth (C^∞) and included in $L_p(\mathbb{R}\times\mathbb{S}^{d-1})$

for all $p \geq 1$. While the members of $\tilde{\mathcal{S}}_{\mathrm{Rad}}$ do not necessarily decay rapidly, the mapping $\mathrm{R}^*\tilde{\mathrm{K}}^*_{\mathrm{rad}} : \tilde{\mathcal{S}}_{\mathrm{Rad}} \to \mathcal{S}(\mathbb{R}^d)$ is still guaranted to be an isomorphism (see [43, Theorem 3.3.1, p. 83] where similar arguments are used). Under the assumption that the hyper-spherical norm $\|\cdot\|_{\mathcal{X}}$ is continuous on $\tilde{\mathcal{S}}_{\mathrm{Rad}}$, we can readily adapt the proof of [57, Theorem 8] to establish the following.

Proposition 3 (Odd Radon-compatible Banach spaces) *Consider the Banach space* $\tilde{\mathcal{X}}_{\mathrm{Rad}} = \overline{(\tilde{\mathcal{S}}_{\mathrm{Rad}}, \|\cdot\|_{\mathcal{X}})}$ *of odd hyper-spherical functions. Then, the following holds.*

1. *The map* $\mathrm{R}^*\tilde{\mathrm{K}}^*_{\mathrm{rad}} : \tilde{\mathcal{X}}_{\mathrm{Rad}} \to \tilde{\mathcal{Y}} = \mathrm{R}^*\tilde{\mathrm{K}}^*_{\mathrm{rad}}\big(\tilde{\mathcal{X}}_{\mathrm{Rad}}\big)$ *is an isometric bijection, with* $\tilde{\mathrm{R}}\mathrm{R}^*\tilde{\mathrm{K}}^*_{\mathrm{rad}} = \mathrm{Id}$ *on* $\tilde{\mathcal{X}}_{\mathrm{Rad}}$.
2. *The map* $\tilde{\mathrm{R}}^* : \tilde{\mathcal{X}}'_{\mathrm{Rad}} \to \tilde{\mathcal{Y}}' = \tilde{\mathrm{R}}^*\big(\tilde{\mathcal{X}}'_{\mathrm{Rad}}\big)$ *is an isometric bijection, with* $\tilde{\mathrm{K}}_{\mathrm{rad}}\mathrm{R}\tilde{\mathrm{R}}^* = \mathrm{Id}$ *on* $\tilde{\mathcal{X}}'_{\mathrm{Rad}}$.

The underlying definition of the "oddified" backprojection operator $\tilde{\mathrm{R}}^* : \tilde{\mathcal{X}}'_{\mathrm{Rad}} \to \tilde{\mathcal{Y}}'$ for $g \in \tilde{\mathcal{X}}'_{\mathrm{Rad}}$ is

$$\langle \tilde{\mathrm{R}}^*\{g\}, \varphi\rangle = \langle g, \tilde{\mathrm{R}}\{\varphi\}\rangle_{\mathrm{Rad}} = \langle \tilde{\mathrm{H}}^*_{\mathrm{rad}}\{g\}, \mathrm{R}\{\varphi\}\rangle_{\mathrm{Rad}} \tag{66}$$

for all $\varphi \in \tilde{\mathcal{Y}}$ or, equivalently, $\varphi \in \mathcal{S}(\mathbb{R}^d)$ since $\mathcal{S}(\mathbb{R}^d)$ is dense in $\mathcal{Y}$ by construction. Likewise, by using the property that the Hilbert transform is a homeomorphism from $\mathcal{S}_{\mathrm{Liz,even}}$ onto $\mathcal{S}_{\mathrm{Liz,odd}}$ [47] with the underlying Lizorkin spaces being included in $\mathcal{S}_{\mathrm{Rad}}$ and $\tilde{\mathcal{S}}_{\mathrm{Rad}} \subset L_{p,\mathrm{odd}}$, respectively, we can adapt the proof of Lemma 2 to show that $\tilde{L}_{q,\mathrm{Rad}} = L_{q,\mathrm{odd}}$ for $q \in (1,\infty)$, $\tilde{C}_{0,\mathrm{Rad}} = C_{0,\mathrm{odd}}$ for $p = \infty$, and $\tilde{\mathcal{M}}_{\mathrm{Rad}} = (\tilde{C}_{0,\mathrm{Rad}})' = \mathcal{M}_{\mathrm{odd}}$.

The bottom line is that the whole argumentation, including the critical Fourier-based proofs of Section 5, applies in this modified setting as well. Accordingly, all theorems that mention the regularization operator $\mathrm{L}_{\mathrm{R}} = \mathrm{K}_{\mathrm{rad}}\mathrm{R}\mathrm{L}$ and the radial profile $\rho_{\mathrm{rad}} = \mathcal{F}^{-1}\{1/\widehat{L}_{\mathrm{rad}}\}$ are also valid for the odd setting where these quantifies are substituted by

$$\tilde{\mathrm{L}}_{\mathrm{R}} \triangleq \tilde{\mathrm{K}}_{\mathrm{rad}}\mathrm{R}\mathrm{L} = \tilde{\mathrm{K}}_{\mathrm{rad}}\mathrm{L}_{\mathrm{rad}}\mathrm{R}, \tag{67}$$

$$\tilde{\rho}_{\mathrm{rad}}(t) \triangleq \mathcal{F}^{-1}\left\{\frac{\mathrm{i}\,\mathrm{sign}(\omega)}{\widehat{L}_{\mathrm{rad}}(\omega)}\right\}(t), \tag{68}$$

where (68) directly follows from (66) and Theorem 4. The conditions for admissibility in Definitions 2 and 3, which are all Fourier-based, remain the same, while the adjusted activation $\tilde{\rho}_{\mathrm{rad}}$ (the 1D Hilbert transform of ρ_{rad}) is real-valued and anti-symmetric because the original Fourier profile $\widehat{L}_{\mathrm{rad}} : \mathbb{R} \to \mathbb{R}$ is symmetric.

We note that our admissibility condition with $\gamma_0 = 1$ for odd variant ($n_0 = 0$) is compatible with the condition used by Barron to prove the universality of neural networks with sigmoidal activations [5]. It is also worth mentioning that the substitution of L_{R} by $\tilde{\mathrm{L}}_{\mathrm{R}}$ has no incidence on the form of the RBF in (53) because of the unitary nature of the Hilbert transform.

$\rho_{\mathrm{rad}}(t)$	$\tilde{\rho}_{\mathrm{rad}}(t)$	$\widehat{L}_{\mathrm{rad}}$ or $\widehat{\tilde{L}}_{\mathrm{rad}}$	n_0 (null space)
Exponential			
$\mathrm{e}^{-\lvert t\rvert}$	N/A	$1+\lvert\omega\rvert^2$	-1 (trivial)
Classical sigmoids			
N/A	$\frac{\tanh(\frac{t}{2})}{2}=\frac{1}{2}+\frac{1}{1+\mathrm{e}^{-t}}$	$\frac{\sinh(\pi\omega)}{\pi}$	0 (bias)
N/A	$\frac{\arctan(t)}{\pi}$	$\omega\mathrm{e}^{\lvert\omega\rvert}$	0
Ridge splines	(of degree $n \in \mathbb{N}$)		
$\frac{1}{2}\lvert t\rvert$ (or ReLU)	$t\log\lvert t\rvert$	$\lvert\omega\rvert^2$	1 (affine)
$\propto t^{2n}\log\lvert t\rvert$	$\frac{\mathrm{sign}(t)\lvert t\rvert^{2n}}{(2n)!}$	$\lvert\omega\rvert^{2n+1}$	$2n \geq 2$ (even)
$\frac{1}{2}\frac{\lvert t\rvert^{2n+1}}{(2n+1)!}$	$\propto t^{2n+1}\log\lvert t\rvert$	$\lvert\omega\rvert^{2n+2}$	$2n+1 \geq 1$ (odd)
Fractional splines	(degree $\alpha \in \mathbb{R}^+\backslash\mathbb{N}$)		
$\frac{\lvert t\rvert^{\alpha}\sin(\frac{\alpha\pi}{2})}{\pi\Gamma(\alpha)}$	$\frac{\mathrm{sign}(t)\lvert t\rvert^{\alpha}\cos(\frac{\alpha\pi}{2})}{\pi\Gamma(\alpha)}$	$\lvert\omega\rvert^{\alpha+1}$	$\lceil\alpha\rceil$

Table 1 Examples of admissible symmetric and anti-symmetric activation functions with their corresponding regularization operator. The anti-symmetric activation is given by (68) and requires the use of the modified filter $\tilde{\mathrm{K}}_{\mathrm{rad}}$ in the statement of Theorem 5.

4.5 Specific Configurations

The proposed framework encompasses a wide variety of kernels and activation functions, with minimal restrictions. For instance, one can start with any strictly positive-definite function $\rho_{\mathrm{rad},0}$ whose Fourier transform is strictly positive, and construct some higher-order variants by (fractional) integration. The variants are such that $\rho_{\mathrm{rad},\gamma_0}(t)=\mathcal{F}^{-1}\{\frac{\widehat{\rho}_{\mathrm{rad},0}(\omega)}{\lvert\omega\rvert^{\gamma_0}}\}(t)$ with suitable $\gamma_0>0$ and are guaranteed to meet the requirements in Definition 3. The simplest scenario is to set $\rho_{\mathrm{rad},0}=\delta$, which maps into a Laplacian-type regularization.

In Table 1, we provide examples of admissible operators together with their corresponding symmetric and anti-symmetric activations. It is noteworthy that the two most popular sigmoids (tanh and arctan) are part of the framework. We can determine the explicit frequency response of their regularization filter and show that it is first-order-admissible with a null space that consists of the polynomials of degree $n_0=0$ (the constants). The symmetric spline activations of odd degree $m-1$ and the anti-symmetric ones of even degree are also known: they coincide with the ridge splines of Parhi and Novak, which are tied to the Radon-domain operator $\mathrm{L}_{\mathrm{rad}}=\frac{\partial^m}{\partial t^m}$ [37]. With the present formulation, we can seamlessly extend this family to fractional orders, in direct analogy with [59], by setting $\mathrm{L}=(-\Delta)^{\frac{\alpha+1}{2}}$.

5 Supporting Mathematical Results

To answer the fundamental issue of the existence of a solution in Theorem 5, we need to 1) prove that the "predual" space $\mathcal{X}_{\mathrm{L}}(\mathbb{R}^d)$ is a proper Banach space (Theorem 9); and 2) establish the weak* continuity of the sampling functionals $\delta(\cdot - \boldsymbol{x}_k)$. As we shall see, both aspects largely rest upon the functional characterization of the Schwartz kernel of the pseudoinverse operator $\mathrm{L}_{\mathrm{R}}^{*\dagger}$ provided in Theorem 7.

5.1 Kernel and Stability of Generalized Inverse Operators

Let $\nu : f \mapsto \langle \nu, f\rangle$ be a linear functional that is acting on some Banach space $\mathcal{X}'$. We recall that ν is weak*-continuous if and only if $\nu \in \mathcal{X}$, where $\mathcal{X}$ is the predual of $\mathcal{X}'$ [44]. The condition that $\mathcal{X}'$ is reflexive (i.e., $\mathcal{X}'' = \mathcal{X}$) is equivalent to the continuity of ν on $\mathcal{X}$, which is the standard condition for analysis. However, when $\mathcal{X}'$ is not reflexive (e.g., $\mathcal{X}' = \mathcal{M} = (C_0)'$), the constraint of weak* continuity is stronger than continuity, contrary to what could be suggested by the qualifier "weak." In that case, the predual space $\mathcal{X} \subseteq \mathcal{X}''$, which is continuously embedded in $\mathcal{X}''$, turns out to be smaller than $\mathcal{X}''$. Therefore, in order to establish the weak* continuity of the Dirac functionals $\delta(\cdot - \boldsymbol{x}_0)$ for the scenarios in Theorem 5, we need to show that $\delta(\cdot - \boldsymbol{x}_0) \in \mathcal{X}_{\mathrm{L}_{\mathrm{R}}}$, which can be reduced to proving that $\mathrm{L}_{\mathrm{R}}^{*\dagger}\{\delta(\cdot - \boldsymbol{x}_0)\} \in \mathcal{X}_{\mathrm{Rad}}$.

Theorem 7 (Properties of the impulse response of $\mathrm{L}_{\mathrm{R}}^{*\dagger} = \mathrm{R}\mathrm{L}_{\mathcal{P}}^{-1*}$) *Let* L *be an isotropic operator such that* $\widehat{L}(\boldsymbol{\omega}) = \widehat{L}_{\mathrm{rad}}(\|\boldsymbol{\omega}\|)$ *where* $\widehat{L}_{\mathrm{rad}} : \mathbb{R} \to \mathbb{R}$ *is a continuous function and* $\rho_{\mathrm{rad}}(t) = \mathcal{F}^{-1}\{1/\widehat{L}_{\mathrm{rad}}\}(t)$. *We consider two cases:*

1. *Trivial null space: If* L *satisfies the admissibility conditions in Definition 2, then* $\mathrm{L}_{\mathrm{R}}^{*\dagger} = \mathrm{R}\mathrm{L}^{-1*}$ *and*
$$\nu_{\boldsymbol{x}_0}(t, \boldsymbol{\xi}) = \mathrm{R}\mathrm{L}^{-1*}\{\delta(\cdot - \boldsymbol{x}_0)\}(t, \boldsymbol{\xi}) = \rho_{\mathrm{rad}}(t - \boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}_0) \tag{69}$$
with $\boldsymbol{x}_0 \in \mathbb{R}^d$ *and* $(t, \boldsymbol{\xi}) \in \mathbb{R} \times \mathbb{S}^{d-1}$. *Moreover,* $\nu_{\boldsymbol{x}_0} \in \mathcal{X}_{\mathrm{Rad}}$ *for* $\mathcal{X} = C_0$ *as well as* $\mathcal{X} = L_q$ *with* $q \in [1, \infty]$.
2. *Nontrivial null space: If* L *satisfies the admissibility conditions in Definition 3 with a polynomial null space of degree* n_0*, then*
$$\begin{aligned}\nu_{\boldsymbol{x}_0}(t, \boldsymbol{\xi}) &= \mathrm{L}_{\mathrm{R}}^{*\dagger}\{\delta(\cdot - \boldsymbol{x}_0)\}(t, \boldsymbol{\xi}) \\ &= \rho_{\mathrm{rad}}(t - \boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}_0) - \sum_{n=0}^{n_0} \frac{(-\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}_0)^n}{n!}\big(\kappa_{\mathrm{rad}} * \partial^n \rho_{\mathrm{rad}}\big)(t)\end{aligned} \tag{70}$$
with $\nu_{\boldsymbol{x}_0} \in \mathcal{X}_{\mathrm{Rad}}$ *for the same spaces as in Item 1, but with* $q \in [2, \infty]$. *Moreover,*
$$\sup_{(\boldsymbol{x}_0, \boldsymbol{\xi}) \in \mathbb{R}^d \times \mathbb{S}^{d-1}} (1 + |\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}_0|)^{-n_0} \|\nu_{\boldsymbol{x}_0}(\cdot, \boldsymbol{\xi})\|_{L_q(\mathbb{R})} < \infty \tag{71}$$
for any $q \in [2, \infty]$.

Proof To show that $\nu_{\boldsymbol{x}_0} \in \mathcal{X}_{\mathrm{Rad}}$, it is sufficient to prove that $\nu_{\boldsymbol{x}_0} \in \mathcal{X}(\mathbb{R}^d \times \mathbb{S}^{d-1})$ since $\nu_{\boldsymbol{x}_0}$ is in the range of the Radon transform by construction.

When the null space of L is trivial, L has a stable convolution inverse so that it suffices to show that $\nu_{\boldsymbol{x}_0} \in L_q(\mathbb{R} \times \mathbb{S}^{d-1}) \cap C_0(\mathbb{R} \times \mathbb{S}^{d-1})$. To that end, we formally identify the isotropic distribution $\rho_{\mathrm{iso}} = \mathrm{L}^{-1*}\{\delta\} = \mathrm{L}^{-1}\{\delta\} = \mathcal{F}^{-1}\{1/\widehat{L}_{\mathrm{rad}}(\|\boldsymbol{\omega}\|)\}$ and apply Proposition 1, which yields

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

Due to our assumptions, this function is such that $\|\nu_{\boldsymbol{x}_0}(\cdot, \boldsymbol{\xi})\|_{L_1} = \|\rho_{\mathrm{rad}}\|_{L_1} < \infty$ for any fixed $\boldsymbol{\xi} \in \mathbb{S}^{d-1}$. Moreover, since $1/\hat{L}_{\mathrm{rad}} \in L_1(\mathbb{R})$, ρ_{rad} is bounded, continuous, and vanishing at infinity (by the Riemann-Lebesgue lemma), which gives $\rho_{\mathrm{rad}}(\cdot - \boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}_0) \in C_0(\mathbb{R})$ for any $\boldsymbol{x}_0 \in \mathbb{R}^d$ and $\boldsymbol{\xi} \in \mathbb{S}^{d-1}$. Since $\rho_{\mathrm{rad}} \in L_\infty(\mathbb{R}) \cap L_1(\mathbb{R})$, we readily deduce that $\rho_{\mathrm{rad}}(\cdot - \boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}_0) \in L_q(\mathbb{R})$ for all intermediate $q \geq 1$, which then also yields $\rho_{\mathrm{rad}}(t - \boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}_0) \in L_q(\mathbb{R} \times \mathbb{S}^{d-1})$ because the surface of the unit sphere $\mathbb{S}^{d-1}$ is bounded. Finally, since $\rho_{\mathrm{rad}} : \mathbb{R} \to \mathbb{R}$ is continuous and vanishing at infinity, $\nu_{\boldsymbol{x}_0}(t, \boldsymbol{\xi})$ is jointly continuous in $(t, \boldsymbol{\xi})$ and vanishing at $t \to \pm\infty$ for all $\boldsymbol{\xi} \in \mathbb{S}^{d-1}$, which implies that $\nu_{\boldsymbol{x}_0} \in C_0(\mathbb{R} \times \mathbb{S}^{d-1})$.

For the more difficult case of a nontrivial null space, we invoke the Fourier-slice theorem to evaluate the 1D Fourier transform of $\nu_{\boldsymbol{x}_0}(\cdot, \boldsymbol{\xi})$ with $\boldsymbol{\xi}$ fixed as

$$\begin{aligned}
\hat{\nu}_{\boldsymbol{x}_0}(\omega, \boldsymbol{\xi}) &= \frac{\mathcal{F}\{\delta(\cdot - \boldsymbol{x}_0) - \sum_{|\boldsymbol{k}| \leq n_0} \langle m_{\boldsymbol{k}}, \delta(\cdot - \boldsymbol{x}_0)\rangle \, m_{\boldsymbol{k}}^*\}(\omega\boldsymbol{\xi})}{\widehat{L}_{\mathrm{rad}}(\omega)} \\
&= \frac{\mathrm{e}^{-\mathrm{i}\omega\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}_0} - \sum_{|\boldsymbol{k}| \leq n_0} \frac{\boldsymbol{x}_0^{\boldsymbol{k}}}{\boldsymbol{k}!} \, \widehat{m}_{\boldsymbol{k}}^*(\omega\boldsymbol{\xi})}{\widehat{L}_{\mathrm{rad}}(\omega)} \\
&= \frac{\mathrm{e}^{-\mathrm{i}\omega\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}_0} - \sum_{|\boldsymbol{k}| \leq n_0} \frac{\boldsymbol{x}_0^{\boldsymbol{k}}}{\boldsymbol{k}!} \, (-\mathrm{i}\omega\boldsymbol{\xi})^{\boldsymbol{k}} \widehat{\kappa}_{\mathrm{rad}}(\omega)}{\widehat{L}_{\mathrm{rad}}(\omega)} \\
&= \frac{\mathrm{e}^{-\mathrm{i}\omega\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}_0} - \sum_{n=0}^{n_0} \frac{(-\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}_0)^n}{n!} (\mathrm{i}\omega)^n \widehat{\kappa}_{\mathrm{rad}}(\omega)}{\widehat{L}_{\mathrm{rad}}(\omega)},
\end{aligned} \tag{72}$$

where the simplification of the summation over $\boldsymbol{k}$ results from the multinomial expansion $(-\mathrm{i}\omega\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}_0)^n = (y_1 + \cdots + y_d)^n = \sum_{|\boldsymbol{k}|=n} \frac{n!}{\boldsymbol{k}!} \boldsymbol{y}^{\boldsymbol{k}}$ with $\boldsymbol{y} = \big(-\mathrm{i}\omega\xi_i x_{0,i}\big)_{i=1}^d$. The delicate point here is that (72) is potentially singular because it has a pole of multiplicity γ_0 at $\omega = 0$. Fortunately, the condition $\gamma_0 \leq n_0 + 1$ ensures that there is a proper pole-zero cancellation: By recalling that $\widehat{\kappa}_{\mathrm{rad}}(\omega) = 1$ for $\omega \in \Omega_0 = [-R_0, R_0]$ with $R_0 = \frac{1}{2}$ and setting $t_0 = \boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}_0$, we identify the numerator as the remainder of the Maclaurin series of $\mathrm{e}^{-\mathrm{i}\omega t_0}$, which is bounded by

$$\begin{aligned}
\left| \mathrm{e}^{-\mathrm{i}\omega t_0} - \sum_{n=0}^{N} \frac{(-\mathrm{i}t_0)^n}{n!} \omega^n \right| &\leq \sup_{\omega \in \mathbb{R}} \left| (-\mathrm{i}t_0)^{N+1} \mathrm{e}^{-\mathrm{i}\omega t_0} \right| \; \frac{|\omega|^{N+1}}{(N+1)!} \\
&\leq \frac{|t_0|^{N+1}}{(N+1)!} \, |\omega|^{N+1}.
\end{aligned} \tag{73}$$

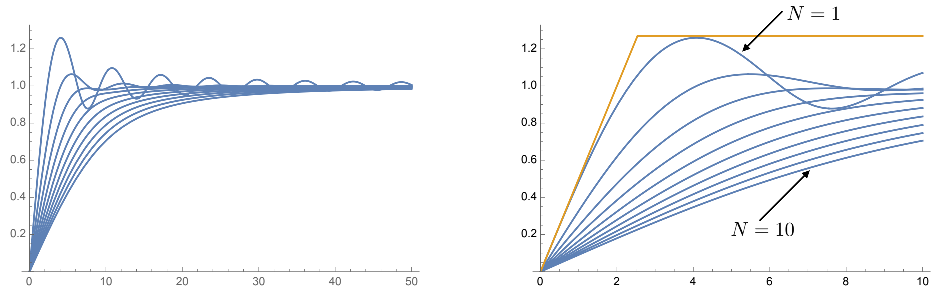

Fig. 1 Functions $\omega \mapsto |r_N(\omega)|$ for $N = 1, \dots, 10$. The dilated plot on the right includes the upper bound specified by (77) as an overlay.

This then yields the estimate

$$|\hat{\nu}_{\boldsymbol{x}_0}(\omega, \boldsymbol{\xi})| \le \frac{|\boldsymbol{\xi}^\mathsf{T}\boldsymbol{x}_0|^{n_0+1}}{C_0(n_0+1)!}|\omega|^{\epsilon} \quad \text{as } \omega \to 0$$

$$\text{with } \epsilon = n_0 + 1 - \gamma_0 = \begin{cases} 0, & \gamma_0 \in \mathbb{N} \\ 1 - (\gamma_0 - \lfloor \gamma_0 \rfloor) \in (0,1), & \text{otherwise.} \end{cases} \tag{74}$$

Since the denominator $\widehat{L}_{\mathrm{rad}}$ in (72) is continuous and non-vanishing away from the origin, $\hat{\nu}_{\boldsymbol{x}_0}(\cdot, \boldsymbol{\xi})$ is bounded on Ω_0, and, by extension, on any compact subset of $\mathbb{R}$. Moreover, since $|\mathrm{e}^{-\mathrm{i}\omega\boldsymbol{\xi}^\mathsf{T}\boldsymbol{x}_0}| = 1$ and $\widehat{\kappa}_{\mathrm{rad}}$ is rapidly decreasing, there exists a constant C such that $|\hat{\nu}_{\boldsymbol{x}_0}(\omega, \boldsymbol{\xi})| \le C|\omega|^{-\gamma_1}$ for $|\omega| > R_1$. This leads to several conclusions. First, if $\gamma_1 > 1$, then $\hat{\nu}_{\boldsymbol{x}_0}(\cdot, \boldsymbol{\xi}) \in L_1(\mathbb{R})$ so that $\nu_{\boldsymbol{x}_0} \in C_{0,\mathrm{Rad}}(\mathbb{R} \times \mathbb{S}^{d-1})$ by the same argument as in the nonsingular case. Second, if $\gamma_1 > \alpha + \frac{1}{2}$, then $\nu_{\boldsymbol{x}_0}(\cdot, \boldsymbol{\xi}) \in W_2^{\alpha}(\mathbb{R})$, where $W_2^{\alpha}(\mathbb{R}) = \{f : \int_{\mathbb{R}}(1 + |\omega|^2)^{\alpha}|\hat{f}(\omega)|^2 \mathrm{d}\omega < \infty\}$ is the Sobolev space of functions with finite-energy derivatives up to order α. Therefore, $\nu_{\boldsymbol{x}_0} \in L_{q,\mathrm{Rad}}(\mathbb{R} \times \mathbb{S}^{d-1})$ for all $q \in [2, \infty]$ provided that $\gamma_1 > 1$. The explicit Radon-domain formula (70) with $\kappa_{\mathrm{rad}}(t) = \mathcal{F}^{-1}\{\widehat{\kappa}_{\mathrm{rad}}(\omega)\}(t)$ is obtained by taking the 1D inverse transform of (72).

To refine our characterization of $\hat{\nu}_{\boldsymbol{x}_0}$, we introduce the function

$$r_N(\omega) = \frac{\mathrm{e}^{-\mathrm{i}\omega} - \sum_{n=0}^{N} \frac{(-\mathrm{i}\omega)^n}{n!}}{\frac{(\mathrm{i}\omega)^N}{N!}} \tag{75}$$

whose modulus is plotted in Figure 1. Some of the remarkable properties of r_N are

$$r_N(\omega) = \frac{-\mathrm{i}\omega}{N+1} \quad \text{as} \quad \omega \to 0 \tag{76}$$

$$\forall \omega \in \mathbb{R}: \quad |r_N(\omega)| \le \min(|\omega|/2, 1.27) \tag{77}$$

$$\lim_{\omega \to \infty} |r_N(\omega)| = 1, \tag{78}$$

with the global bound (77) being overlaid on the graph to demonstrate its sharpness. This function will allows us to control the behavior of

$$\begin{aligned}\hat{\nu}_{\boldsymbol{x}_0}(\omega,\boldsymbol{\xi}) &= \frac{\mathrm{e}^{-\mathrm{i}t_0\omega} - \sum_{n\le n_0} \frac{(-\mathrm{i}t_0\omega)^n}{n!}\kappa_{\mathrm{rad}}(\omega)}{\widehat{L}_{\mathrm{rad}}(\omega)}\\ &= \frac{\widehat{\kappa}_{\mathrm{rad}}(\omega) r_{n_0}(t_0\omega)\frac{(-\mathrm{i}t_0\omega)^{n_0}}{n_0!} + \big(1-\widehat{\kappa}_{\mathrm{rad}}(\omega)\big)\mathrm{e}^{-\mathrm{i}t_0\omega}}{\widehat{L}_{\mathrm{rad}}(\omega)}\end{aligned}$$

by splitting the frequency domain in three regions. First, for $\omega \in \Omega_0 = [0, R_0]$, where $\widehat{\kappa}_{\mathrm{rad}}(\omega) = 1$, we find that

$$|\hat{\nu}_{\boldsymbol{x}_0}(\omega,\boldsymbol{\xi})| = \left|\frac{r_{n_0}(t_0\omega)\frac{(-\mathrm{i}t_0\omega)^{n_0}}{n_0!}}{\widehat{L}_{\mathrm{rad}}(\omega)}\right| \le |t_0|^{n_0}\min\left(|\omega|, 2\right)\frac{2\frac{|\omega|^{n_0}}{n_0!}}{|\widehat{L}_{\mathrm{rad}}(\omega)|}. \tag{79}$$

For the transition region $\omega \in \Omega_{01} = [R_0, \tilde{R}_1]$ with $\tilde{R}_1 = \max(2R_0, R_1)$, where $0 \le \widehat{\kappa}_{\mathrm{rad}}(\omega) \le 1$, we bound the numerator by its maximum, which yields

$$|\hat{\nu}_{\boldsymbol{x}_0}(\omega,\boldsymbol{\xi})| \le \frac{2\frac{|t_0\tilde{R}_1|^{n_0}}{n_0!} + 1}{|\widehat{L}_{\mathrm{rad}}(\omega)|}. \tag{80}$$

Finally, for $\omega \in \Omega_1 = [\tilde{R}_1, \infty]$, where $\widehat{\kappa}_{\mathrm{rad}}(\omega) = 0$, we get the expected tail behavior

$$|\hat{\nu}_{\boldsymbol{x}_0}(\omega,\boldsymbol{\xi})| \le \frac{1}{|\widehat{L}_{\mathrm{rad}}(\omega)|} \le \frac{1}{C_1|\omega|^{\gamma_1}}. \tag{81}$$

Based on those bounds with $t_0 = 1$, we define the auxiliary function

$$u(\omega) = \begin{cases}\min\left(|\omega|, 2\right)\frac{2\frac{|\omega|^{n_0}}{n_0!}}{|\widehat{L}_{\mathrm{rad}}(\omega)|}, & |\omega| < R_0\\ \frac{2\frac{\tilde{R}_1^{n_0}}{n_0!}+1}{|\widehat{L}_{\mathrm{rad}}(\omega)|}, & R_0 \le |\omega| \le \tilde{R}_1\\ \frac{1}{|\widehat{L}_{\mathrm{rad}}(\omega)|}, & \tilde{R}_1 < |\omega|\end{cases} \tag{82}$$

which, by construction, is such that $\|u\|_{L_p} < \infty$ for any $p \ge 1$. We can now use (79), (80), and (81) to bound the norm of $\hat{\nu}_{\boldsymbol{x}_0}(\cdot,\boldsymbol{\xi})$ by distinguishing between two cases. For $t_0 \le 1$, we have $\|\hat{\nu}_{\boldsymbol{x}_0}(\cdot,\boldsymbol{\xi})\|_{L_p} \le \|u\|_{L_p}$, while, for $t_0 \ge 1$, we get $\|\hat{\nu}_{\boldsymbol{x}_0}(\cdot,\boldsymbol{\xi})\|_{L_p} \le |t_0|^{n_0}\|u\|_{L_p}$. By combining these two inequalities, we obtain the universal norm estimate

$$\|\hat{\nu}_{\boldsymbol{x}_0}(\cdot,\boldsymbol{\xi})\|_{L_p(\mathbb{R})} \le (1 + |\boldsymbol{\xi}^\mathsf{T}\boldsymbol{x}_0|)^{n_0}\|u\|_{L_p} < +\infty, \tag{83}$$

which holds for all $(\boldsymbol{x}_0, \boldsymbol{\xi}) \in \mathbb{R}^d \times \mathbb{S}^{d-1}$ and $p \ge 1$. Finally, we invoke the boundedness of the (inverse) Fourier transform $\mathcal{F}^{-1} : L_p(\mathbb{R}) \to L_q(\mathbb{R})$ for $p \in [1, 2]$ to obtain the generic bound

$$\sup_{(\boldsymbol{x}_0,\boldsymbol{\xi})\in\mathbb{R}^d\times\mathbb{S}^{d-1}} (1 + |\boldsymbol{\xi}^\mathsf{T}\boldsymbol{x}_0|)^{-n_0}\|\nu_{\boldsymbol{x}_0}(\cdot,\boldsymbol{\xi})\|_{L_q(\mathbb{R})} < \infty, \tag{84}$$

which is the desired result. $\square$

As complement to the proof of Theorem 7, we make the following remarks.

1. The function $\omega \mapsto \hat{\nu}_{\boldsymbol{x}_0}(\omega, \boldsymbol{\xi})$ in (72) is ϵ-Hölder continuous around the origin for $\gamma_0 \notin 2\mathbb{N}$ and as smooth as $\widehat{L}$ when $\gamma_0 = 2n$. For instance, the case where $\mathrm{L} = (-\Delta)^n$ is a non-fractional iterate of the Laplacian corresponds to $\widehat{L}_{\mathrm{rad}}(\omega) = \omega^{2n}$ and $\hat{\nu}_{\boldsymbol{x}_0}(\cdot, \boldsymbol{\xi}) \in C^{\infty}(\mathbb{R})$, which then translates into $t \mapsto \nu_{\boldsymbol{x}_0}(t, \boldsymbol{\xi})$ being rapidly decreasing, but with a limited order of differentiability controlled by $\gamma_0 = \gamma_1 = 2n$.

2. The proof can be readily adapted to characterize the partial derivatives of $\nu_{\boldsymbol{x}_0}$ by replacing $\delta(\cdot - \boldsymbol{x}_0)$ by $\partial^{\boldsymbol{n}}\delta(\cdot - \boldsymbol{x}_0)$ with $|\boldsymbol{n}| < n_0$. These distributions are such that $\langle \partial^{\boldsymbol{n}}\delta(\cdot - \boldsymbol{x}_0), m_{\boldsymbol{k}} \rangle = (-1)^{|\boldsymbol{n}|}\partial^{\boldsymbol{n}} m_{\boldsymbol{k}}(\boldsymbol{x}_0) = (-1)^{|\boldsymbol{n}|} m_{\boldsymbol{k}-\boldsymbol{n}}(\boldsymbol{x}_0)$, with the convention that $m_{\boldsymbol{k}-\boldsymbol{n}} = 0$ if $k_m < n_m$ for any $m \in \{1, \dots, d\}$.

3. The leading term in (70) is $\rho_{\mathrm{rad}}(t - \tau_0)$ with $\tau_0 = \boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}_0$, which, in the non-trivial scenario, typically grows as $O(|t|^{\gamma_0 - 1})$. Our analysis shows that the second correction term in (70), which is unbounded at infinity as well, essentially neutralizes this growth. It is tempting to call this a miraculous cancellation.

The bottom line is that, for any $(\boldsymbol{x}_0, \boldsymbol{\xi}_0) \in \mathbb{R}^d \times \mathbb{S}^{d-1}$, the function $t \mapsto \nu_{\boldsymbol{x}_0}(t, \boldsymbol{\xi}_0)$ is continuous, bounded with a maximum that is proportional to $|\boldsymbol{\xi}_0^{\mathsf{T}}\boldsymbol{x}_0|^{n_0}$ (see (84) with $p = \infty$), and vanishing at infinity. This is a remarkable property that also guarantees the boundedness of $\mathrm{L}_{\mathrm{R}}^{*\dagger} : L_{1,n_0}(\mathbb{R}^d) \to \mathcal{X}_{\mathrm{Rad}}$, which is not obvious a priori. The enabling ingredient there is (71), which ensures that the corresponding bounding constant in Theorem 8 is finite. Indeed, since $\|\boldsymbol{x}\| \geq |\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}|$ with equality if and only if $\boldsymbol{\xi}$ and $\boldsymbol{x}$ are colinear, we have that

$$\begin{aligned}\|\mathrm{L}_{\mathrm{R}}^{*\dagger}\| &\leq \sup_{\boldsymbol{x} \in \mathbb{R}^d,\, \boldsymbol{\xi} \in \mathbb{S}^{d-1}} (1 + \|\boldsymbol{x}\|)^{-n_0} \|\nu_{\boldsymbol{x}}(\cdot, \boldsymbol{\xi})\|_{\mathcal{X}(\mathbb{R})} \\ &\leq \sup_{\boldsymbol{x} \in \mathbb{R}^d,\, \boldsymbol{\xi} \in \mathbb{S}^{d-1}} (1 + |\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{x}|)^{-n_0} \|\nu_{\boldsymbol{x}}(\cdot, \boldsymbol{\xi})\|_{\mathcal{X}(\mathbb{R})} < \infty.\end{aligned}$$

Theorem 8 (Stability of Cartesian-to-Radon-domain mappings) *Let $h_{\boldsymbol{x}}(t, \boldsymbol{\xi}) = \mathrm{T}\{\delta(\cdot - \boldsymbol{x})\}(t, \boldsymbol{\xi})\}$ denote the generalized impulse response of the operator $\mathrm{T} : L_{1,\alpha}(\mathbb{R}^d) \to \mathcal{X}_{\mathrm{Rad}}(\mathbb{R} \times \mathbb{S}^{d-1})$ and let*

$$C_{p,\alpha} = \sup_{\boldsymbol{x} \in \mathbb{R}^d,\, \boldsymbol{\xi} \in \mathbb{S}^{d-1}} (1 + \|\boldsymbol{x}\|)^{-\alpha} \|h_{\boldsymbol{x}}(\cdot, \boldsymbol{\xi})\|_{L_p(\mathbb{R})}. \tag{85}$$

1. *$\mathcal{X} = C_0$: If $C_{\infty,\alpha} < \infty$, then $\mathrm{T} : L_{1,\alpha}(\mathbb{R}^d) \to C_{0,\mathrm{Rad}}$ is bounded with $\|\mathrm{T}\| \leq C_{\infty,\alpha}$.*
2. *$\mathcal{X} = L_p$ with $p \in (1, \infty)$: If $C_{p,\alpha} < \infty$, then $\mathrm{T} : L_{1,\alpha}(\mathbb{R}^d) \to L_{p,\mathrm{Rad}}$ is bounded with $\|\mathrm{T}\| \leq \frac{2\pi^{d/2}}{\Gamma(d/2)} C_{p,\alpha}$.*

The same holds true for the adjoint $\mathrm{T}^* : \mathcal{X}'_{\mathrm{Rad}} \to L_{\infty,-\alpha}(\mathbb{R}^d)$ with $\|\mathrm{T}^*\| = \|\mathrm{T}\|$.

Proof The function $\big((t,\boldsymbol{\xi}),\boldsymbol{x}\big)\mapsto h_{\boldsymbol{x}}(t,\boldsymbol{\xi})$ is the Schwartz kernel of the operator T, so that

$$\begin{aligned}|g(t,\boldsymbol{\xi})| &= \big|\mathrm{T}\{f\}(t,\boldsymbol{\xi})\big| = \Big|\int_{\mathbb{R}^d} h_{\boldsymbol{x}}(t,\boldsymbol{\xi})\ f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\Big|\\ &\le \sup_{\boldsymbol{x}\in\mathbb{R}^d,\ \boldsymbol{\xi}\in\mathbb{S}^{d-1}} \big((1+\|\boldsymbol{x}\|)^{-\alpha}\|h_{\boldsymbol{x}}(\cdot,\boldsymbol{\xi})\|_{L_\infty}\big)\ \int_{\mathbb{R}^d}(1+\|\boldsymbol{x}\|)^{\alpha}|f(\boldsymbol{x})|\mathrm{d}\boldsymbol{x}.\end{aligned}$$

Consequently,

$$\|g\|_{L_\infty} = \sup_{(t,\boldsymbol{\xi})\in\mathbb{R}\times\mathbb{S}^{d-1}} |g(t,\boldsymbol{\xi})| = C_{\infty,\alpha}\|f\|_{L_{1,\alpha}},$$

which is the desired bound for $\mathcal{X}=C_0$.

To handle the reflexive case $\mathcal{X}=L_p$, we consider the adjoint operator T^* whose Schwartz kernel $\big(\boldsymbol{x},(t,\boldsymbol{\xi})\big)\mapsto h_{\boldsymbol{x}}(t,\boldsymbol{\xi})$ is obtained by transposition. We now show that $\mathrm{T}^*: L_q(\mathbb{R}\times\mathbb{S}^{d-1})\to L_{\infty,-\alpha}(\mathbb{R}^d)$ with $q=\frac{p}{p-1}$ is bounded which, by duality, implies that the same holds true for $(\mathrm{T}^*)^*=\mathrm{T}: L_{1,\alpha}(\mathbb{R}^d)\xrightarrow{\text{c.}} L_p(\mathbb{R}\times\mathbb{S}^{d-1})$ since $L_{1,\alpha}(\mathbb{R}^d)$ is isometrically embedded in its bidual $\big(L_{1,\alpha}(\mathbb{R}^d)\big)''=\big(L_{\infty,-\alpha}(\mathbb{R}^d)\big)'$. The action of the adjoint operator is described as

$$f(\boldsymbol{x})=\mathrm{T}^*\{g\}(\boldsymbol{x})=\int_{\mathbb{R}}\int_{\mathbb{S}^{d-1}} h_{\boldsymbol{x}}(t,\boldsymbol{\xi})\ g(t,\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}\mathrm{d}t$$

which, with the help of Hölder's inequality, yields the bound

$$\begin{aligned}\big|(1+\|\boldsymbol{x}\|)^{-\alpha}f(\boldsymbol{x})\big| &\le (1+\|\boldsymbol{x}\|)^{-\alpha}\|h_{\boldsymbol{x}}\|_{L_p(\mathbb{R}\times\mathbb{S}^{d-1})}\ \|g\|_{L_q(\mathbb{R}\times\mathbb{S}^{d-1})}\\ &\le \sup_{\boldsymbol{x}\in\mathbb{R}^d,\ \boldsymbol{\xi}\in\mathbb{S}^{d-1}}\big((1+\|\boldsymbol{x}\|)^{-\alpha}S_d\ \|h_{\boldsymbol{x}}(\cdot,\boldsymbol{\xi})\|_{L_p(\mathbb{R})}\big)\ \|g\|_{L_q(\mathbb{R}\times\mathbb{S}^{d-1})}\\ &\le S_d\ C_{p,\alpha}\ \|g\|_{L_q(\mathbb{R}\times\mathbb{S}^{d-1})},\end{aligned}$$

where $S_d=\frac{2\pi^{d/2}}{\Gamma(d/2)}$ is the surface of the unit hypersphere $\mathbb{S}^{d-1}$. This proves the desired result with $\|\mathrm{T}\|=\|\mathrm{T}^*\|\le S_dC_{p,\alpha}$. □

5.2 Characterization of the Predual Space $\mathcal{X}_{\mathrm{L_R}}$

The application of the general representer in [58] requires that the native space $\mathcal{X}'_{\mathrm{L_R}}=(\mathcal{X}_{\mathrm{L_R}})'$ be identifiable as the topological dual of some primary Banach space $\mathcal{X}_{\mathrm{L_R}}$. The specification of the proper predual space is achieved constructively through a completion process that ensures that $\mathcal{X}_{\mathrm{L_R}}$ is a complete normed space (Banach property).

Theorem 9 (Construction of the predual Banach space $\mathcal{X}_{\mathrm{L_R}}$) *Let* L *be an isotropic LSI operator with radial frequency profile* $\widehat{L}_{\mathrm{rad}}$ *that is spline-admissible with a polynomial null space* $\mathcal{P}$ *of degree* n_0. *Then,* $\mathrm{L}_{\mathrm{R}}^{*\dagger}=\mathrm{R}\mathrm{L}^{-1*}(\mathrm{Id}-$

$\mathrm{Proj}_{\mathcal{P}'}) : L_{1,n_0}(\mathbb{R}^d) \xrightarrow{\text{c.}} \mathcal{X}_{\mathrm{Rad}}$ *is bounded, and admits the unique extension* $\mathrm{L}_{\mathrm{R}}^{*\dagger} : \mathcal{X}_{\mathrm{L_R}}(\mathbb{R}^d) \xrightarrow{\text{c.}} \mathcal{X}_{\mathrm{Rad}}$ *with the properties*

$$\forall v \in \mathcal{X}_{\mathrm{Rad}} : \quad \mathrm{L}_{\mathrm{R}}^{*\dagger}\mathrm{L}_{\mathrm{R}}^{*}\{v\} = v \tag{86}$$

$$\forall p_0^* \in \mathcal{P}' : \quad \mathrm{L}_{\mathrm{R}}^{*\dagger}\{p_0^*\} = 0 \tag{87}$$

$$\forall f \in \mathcal{X}_{\mathrm{L_R}}(\mathbb{R}^d) : \quad \mathrm{L}_{\mathrm{R}}^{*}\mathrm{L}_{\mathrm{R}}^{*\dagger}\{f\} = (\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}'})\{f\} = \mathrm{Proj}_{\mathcal{U}}\{f\}, \tag{88}$$

where $\mathrm{L}_{\mathrm{R}}^{*} = \mathrm{L}^*\mathrm{R}^*\mathrm{K}_{\mathrm{rad}}$ *and* $\mathcal{X}_{\mathrm{L_R}}(\mathbb{R}^d) = \mathrm{L}_{\mathrm{R}}^{*}(\mathcal{X}_{\mathrm{Rad}}) \oplus \mathcal{P}'$ *is equipped with the norm* $\|f\|_{\mathcal{X}_{\mathrm{L_R}}} \triangleq \max(\|\mathrm{L}_{\mathrm{R}}^{*\dagger}\{f\}\|_{\mathcal{X}}, \|\mathrm{Proj}_{\mathcal{P}'}\{f\}\|_{\mathcal{P}'})$. *The space* $\mathcal{X}_{\mathrm{L_R}}$ *is complete and isomorphic to* $\mathcal{X}_{\mathrm{Rad}} \times \mathcal{P}'$ *with* $f = \mathrm{L}_{\mathrm{R}}^{*}\{v\} + p_0^* \mapsto (v, p_0^*) = (\mathrm{L}_{\mathrm{R}}^{*\dagger}\{f\}, \mathrm{Proj}_{\mathcal{P}'}\{f\})$. *Moreover,* $\mathcal{S}(\mathbb{R}^d) \overset{\text{d.}}{\hookrightarrow} L_{1,n_0}(\mathbb{R}^d) \overset{\text{d.}}{\hookrightarrow} \mathcal{X}_{\mathrm{L_R}}(\mathbb{R}^d) \overset{\text{d.}}{\hookrightarrow} \mathcal{S}'(\mathbb{R}^d)$ *with the embedding being continuous and dense.*

Theorem 9 obviously also applies to scenarios where the null space is trivial by setting $\mathcal{P}' = \{0\}$ and only retaining (86), in which case $\mathcal{X}_{\mathrm{L_R}}(\mathbb{R}^d) = \mathrm{L}_{\mathrm{R}}^{*}(\mathcal{X}_{\mathrm{Rad}})$ and $\mathcal{S}(\mathbb{R}^d) \overset{\text{d.}}{\hookrightarrow} L_1(\mathbb{R}^d) \overset{\text{d.}}{\hookrightarrow} \mathcal{X}_{\mathrm{L_R}}(\mathbb{R}^d) \overset{\text{d.}}{\hookrightarrow} \mathcal{S}'(\mathbb{R}^d)$.

Proof Since $\mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\big(\mathcal{S}_{\mathrm{Rad}}\big) = \mathcal{S}(\mathbb{R}^d)$, the image of $\mathcal{S}_{\mathrm{Rad}}$ under $\mathrm{L}_{\mathrm{R}}^{*}$ is a vector space denoted by $\mathcal{S}_{\mathrm{L}_{\mathrm{R}}^{*}}(\mathbb{R}^d) = \mathrm{L}_{\mathrm{R}}^{*}\big(\mathcal{S}_{\mathrm{Rad}}\big) = \mathrm{L}^*\big(\mathcal{S}(\mathbb{R}^d)\big)$. The spline-admissibility of L implies that L^* is injective on $\mathcal{S}(\mathbb{R}^d)$ which, in turn, translates into the injectivity of $\mathrm{L}_{\mathrm{R}}^{*}$ on $\mathcal{S}_{\mathrm{Rad}}$. The latter statement is equivalent to the existence of a linear operator $\mathrm{L}_{\mathrm{R}}^{*-1}\big|_{\mathcal{S}_{\mathrm{L}_{\mathrm{R}}^{*}}} = \mathrm{L}_{\mathrm{R}}^{*-1}$ (for short) such that, for any $u = \mathrm{L}_{\mathrm{R}}^{*}\{\phi\} \in \mathcal{S}_{\mathrm{L}_{\mathrm{R}}^{*}}(\mathbb{R}^d)$ with $\phi \in \mathcal{S}_{\mathrm{Rad}}$, it holds that

$$\mathrm{L}_{\mathrm{R}}^{*-1}\{u\} = \mathrm{L}_{\mathrm{R}}^{*-1}\mathrm{L}_{\mathrm{R}}^{*}\{\phi\} = \phi.$$

Hence, if $\phi \mapsto \|\phi\|_{\mathcal{X}}$ is a norm on $\mathcal{S}_{\mathrm{Rad}}$, then the same holds true for $u \mapsto \|u\|_{\mathcal{U}} \triangleq \|\mathrm{L}_{\mathrm{R}}^{*-1}\{u\}\|_{\mathcal{X}}$ on $\mathcal{S}_{\mathrm{L}_{\mathrm{R}}^{*}}(\mathbb{R}^d)$. This means that the normed spaces $(\mathcal{S}_{\mathrm{Rad}}, \|\cdot\|_{\mathcal{X}})$ and $(\mathcal{S}_{\mathrm{L}_{\mathrm{R}}^{*}}(\mathbb{R}^d), \|\cdot\|_{\mathcal{U}})$ are (isometrically) isomorphic. Moreover, since $\mathcal{S}_{\mathrm{L}_{\mathrm{R}}^{*}}(\mathbb{R}^d) \subset L_{1,n_0}(\mathbb{R}^d)$ (Condition 4 in Definition 3) and $\mathcal{S}_{\mathrm{Rad}} \overset{\text{d.}}{\hookrightarrow} \mathcal{X}_{\mathrm{Rad}}$, the inverse operator $\mathrm{L}_{\mathrm{R}}^{*-1}\big|_{\mathcal{S}_{\mathrm{L}_{\mathrm{R}}^{*}}}$ coincides on $\mathcal{S}_{\mathrm{L}_{\mathrm{R}}^{*}}$ with $\mathrm{L}_{\mathrm{R}}^{*\dagger} : L_{1,n_0}(\mathbb{R}^d) \xrightarrow{\text{c.}} \mathcal{X}_{\mathrm{Rad}}$ whose impulse response is characterized in Theorem 7. The well-posedness and boundedness of $\mathrm{L}_{\mathrm{R}}^{*\dagger}$ on $L_{1,n_0}(\mathbb{R}^d)$ for $n_0 = \lceil \gamma_0 - 1 \rceil$ follows from Theorem 8 and (71) in Theorem 7, which provides the required stability condition. The other fundamental ingredient is $\langle m_{\boldsymbol{k}}, u\rangle = \langle m_{\boldsymbol{k}}, \mathrm{L}_{\mathrm{R}}^{*}\{\phi\}\rangle = \langle \mathrm{L}_{\mathrm{R}}\{m_{\boldsymbol{k}}\}, \phi\rangle_{\mathrm{Rad}} = 0$ for all $u \in \mathcal{S}_{\mathrm{L}^*}(\mathbb{R}^d)$ and $|\boldsymbol{k}| \le n_0$, which implies that $\mathrm{Proj}_{\mathcal{P}'}\{u\} = 0$. Consequently, we have that

$$\mathrm{L}_{\mathrm{R}}^{*\dagger}\{u\} = \mathrm{L}_{\mathrm{R}}^{*-1}(\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}'})\{u\} = \mathrm{L}_{\mathrm{R}}^{*-1}\{u\} = \mathrm{L}_{\mathrm{R}}^{*-1}\mathrm{L}_{\mathrm{R}}^{*}\{\phi\} = \phi,$$

which confirms the equivalence of $\mathrm{L}_{\mathrm{R}}^{*\dagger}$ and $\mathrm{L}_{\mathrm{R}}^{*-1}\big|_{\mathcal{S}_{\mathrm{L}_{\mathrm{R}}^{*}}}$.

So far, we have shown that the operator $\mathrm{L}_{\mathrm{R}}^{*\dagger} : \big(\mathcal{S}_{\mathrm{L}_{\mathrm{R}}^{*}}, \|\cdot\|_{\mathcal{U}}\big) \to \mathcal{X}_{\mathrm{Rad}}$ is an isometry. As next step, we invoke the bounded linear transformation (BLT) theorem [44, Theorem I.7, p 9] to uniquely extend the operator to the completed space $\mathcal{U} = \overline{(\mathcal{S}_{\mathrm{L}_{\mathrm{R}}^*}, \|\cdot\|_{\mathcal{U}})}$ which, by construction, is the Banach space equipped with the norm $\|\cdot\|_{\mathcal{U}}$. This extension argument also applies the other way around: Since $(\mathcal{S}_{\mathrm{L}^*}(\mathbb{R}^d), \|\cdot\|_{\mathcal{U}}) \xhookrightarrow{\text{iso.}} \mathcal{U}$, the operator $\mathrm{L}_{\mathrm{R}}^* : (\mathcal{S}_{\mathrm{Rad}}, \|\cdot\|_{\mathcal{X}}) \to \mathcal{U}$ has a unique (isometric) extension $\mathrm{L}_{\mathrm{R}}^* : \mathcal{X}_{\mathrm{Rad}} \to \mathcal{U}$ with $\mathcal{X}_{\mathrm{Rad}}$ being the closure $\mathcal{X}_{\mathrm{Rad}} = \overline{(\mathcal{S}_{\mathrm{Rad}}, \|\cdot\|_{\mathcal{X}})}$. This proves that the spaces $\mathcal{X}_{\mathrm{Rad}}$ and $\mathcal{U}$ are isometrically isomorphic with $\mathcal{U} = \mathrm{L}_{\mathrm{R}}^*(\mathcal{X}_{\mathrm{Rad}})$ and $\mathcal{X}_{\mathrm{Rad}} = \mathrm{L}_{\mathrm{R}}^{*\dagger}(\mathcal{U})$. In addition, we have that $\mathcal{U} \perp \mathcal{P}$, which means that $\mathrm{Proj}_{\mathcal{P}'}\{u\} = 0$ for all $u \in \mathcal{U}$. Since $\mathcal{U}$ and $\mathcal{P}'$ are both Banach spaces, the direct-sum space $\mathcal{X}_{\mathrm{L}_{\mathrm{R}}} = \mathcal{U} \oplus \mathcal{P}'$, equipped with the composite norm $\|f\|_{\mathcal{X}_{\mathrm{L}_{\mathrm{R}}}} = \max(\|\mathrm{Proj}_{\mathcal{U}}\{f\}\|_{\mathcal{U}}, \|\mathrm{Proj}_{\mathcal{P}'}\{f\}\|_{\mathcal{P}'})$, is complete (Banach property) and isomorphic to $\mathcal{X}_{\mathrm{Rad}} \times \mathcal{P}'$. The final element is to recognize that $\|\mathrm{Proj}_{\mathcal{U}}\{f\}\|_{\mathcal{U}} = \|\mathrm{L}_{\mathrm{R}}^{*\dagger}\mathrm{Proj}_{\mathcal{U}}\{f\}\|_{\mathcal{X}} = \|\mathrm{L}_{\mathrm{R}}^{*\dagger}\{f\}\|_{\mathcal{X}}$, where $\mathrm{Proj}_{\mathcal{U}} = (\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}'})$. This direct-sum decomposition has an equivalent description in terms of operators, which is the form given in the statement of Theorem 9. Specifically, the isomorphism between $\mathcal{U}$ and $\mathcal{X}$ is expressed by (86) and (88) for $f \in \mathcal{U}$. This is complemented by the null-space property (87), which ensures that the components of f that are in $\mathcal{P}'$ are annihilated by $\mathrm{L}_{\mathrm{R}}^*$.

Embeddings: The denseness of $\mathcal{S}$ in $\mathcal{X}_{\mathrm{L}_{\mathrm{R}}}$ follows from the observation that $\mathcal{S}(\mathbb{R}^d) = (\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}'})(\mathcal{S}(\mathbb{R}^d)) \oplus \mathcal{P}'$. Since, by construction, one has that $(\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}'})(\mathcal{S}) \xhookrightarrow{\text{d.}} \mathcal{U}$ and $(\mathrm{Id} - \mathrm{Proj}_{\mathcal{P}'})(L_{1,n_0}) \xhookrightarrow{\text{d.}} \mathcal{U}$, one also has that $\mathcal{S}(\mathbb{R}^d) \xhookrightarrow{\text{d.}} L_{1,n_0}(\mathbb{R}^d) \xhookrightarrow{\text{d.}} \mathcal{X}_{\mathrm{L}_{\mathrm{R}}}(\mathbb{R}^d)$.

As for the relation $\mathcal{X}_{\mathrm{L}_{\mathrm{R}}} = \mathcal{U} \oplus \mathcal{P}' \xhookrightarrow{\text{d.}} \mathcal{S}'(\mathbb{R}^d)$, we already have that $\mathcal{P}' \hookrightarrow \mathcal{S}(\mathbb{R}^d) \hookrightarrow \mathcal{S}'(\mathbb{R}^d)$, by construction. To show that $\mathcal{U} \hookrightarrow \mathcal{S}'(\mathbb{R}^d)$, we invoke the intertwining relation and express $\mathcal{U} = \mathrm{L}^*\mathrm{R}^*\mathrm{K}_{\mathrm{rad}}(\mathcal{X}_{\mathrm{Rad}})$ as $\mathcal{U} = \mathrm{R}^*\mathrm{Q}_{\mathrm{rad}}(\mathcal{X}_{\mathrm{Rad}})$, where $\mathrm{Q}_{\mathrm{rad}} = \mathrm{L}_{\mathrm{rad}}\mathrm{K}_{\mathrm{rad}} : \mathcal{X}_{\mathrm{Rad}} \to \mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1})$. Our hypotheses on $\widehat{L}_{\mathrm{rad}}$ ensure that $\mathrm{Q}_{\mathrm{rad}}\{\phi\}$ is well-defined for every $\phi \in \mathcal{X}$, with the operator being continuous in the weak topology of $\mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1})$. (As the latter is a complete nuclear space, the weak (sequential) convergence also ensures continuity in the strong topology [55].) Since $\mathrm{R}^* : \mathcal{S}'(\mathbb{R} \times \mathbb{S}^{d-1}) \xrightarrow{\text{c.}} \mathcal{S}'(\mathbb{R}^d)$, as implied by (27), the composed operator $\mathrm{R}^*\mathrm{Q}_{\mathrm{rad}} : \mathcal{X}_{\mathrm{Rad}} \to \mathcal{S}'(\mathbb{R}^d)$ is continuous as well, which proves that $\mathcal{U}$ and $\mathcal{X}_{\mathrm{L}_{\mathrm{R}}}$ are both continuously embedded in $\mathcal{S}'(\mathbb{R}^d)$. The latter embedding is also dense by transitivity since $\mathcal{S}(\mathbb{R}^d) \subset \mathcal{X}_{\mathrm{L}_{\mathrm{R}}}$ and $\mathcal{S}(\mathbb{R}^d) \xhookrightarrow{\text{d.}} \mathcal{S}'(\mathbb{R}^d)$ [51,19].

□

A Appendix: Duality Maps

The generalization of the Cauchy-Schwarz inequality for any dual pair $(\mathcal{X}, \mathcal{X}')$ of Banach spaces is the so-called *duality inequality*

$$\forall (f, x) \in \mathcal{X}' \times \mathcal{X} : \langle f, x\rangle_{\mathcal{X}' \times \mathcal{X}} \le \|f\|_{\mathcal{X}'}\|x\|_{\mathcal{X}}, \tag{89}$$

which is tightly linked to the definition of the dual norm given by

$$\|f\|_{\mathcal{X}'} \triangleq \sup_{x \in \mathcal{X} : \|x\|_{\mathcal{X}} \le 1} \langle f, x\rangle_{\mathcal{X}' \times \mathcal{X}}. \tag{90}$$

Definition 8 (Strict convexity) A Banach space $\mathcal{X}$ is said to be strictly convex if, for all $x_1, x_2 \in \mathcal{X}$ such that $\|x_1\|_{\mathcal{X}} = \|x_2\|_{\mathcal{X}} = 1$ and $x_1 \neq x_2$, one has that $\|\theta x_1 + (1-\theta)x_2\|_{\mathcal{X}} < 1$ for all $\theta \in (0, 1)$.

It is obvious from (90) that (89) is sharp. Moreover, when $\mathcal{X}$ is reflexive and strictly convex, there is a single element $x^* \in \mathcal{X}'$ (the Banach conjugate of x) such that $\|x^*\|_{\mathcal{X}'} = \|x\|_{\mathcal{X}}$ (isometry) and $\langle x^*, x\rangle_{\mathcal{X}'\times\mathcal{X}} = \|x^*\|_{\mathcal{X}'}\|x\|_{\mathcal{X}}$ (sharp duality bound) [15]. This leads to the definition of the corresponding duality map $\mathrm{J}_{\mathcal{X}} : \mathcal{X} \to \mathcal{X}'$ as

$$\mathrm{J}_{\mathcal{X}}\{x\} = x^*. \tag{91}$$

Since the dual of $\mathcal{X}'$ is strictly convex as well, we have that $\mathrm{J}_{\mathcal{X}}^{-1} = \mathrm{J}_{\mathcal{X}'} : \mathcal{X}' \to \mathcal{X}$ with $\mathrm{J}_{\mathcal{X}'}\{x^*\} = x$, where $(x^*)^* = x \in \mathcal{X}'' = \mathcal{X}$ is the unique Banach conjugate of $x^* \in \mathcal{X}'$.

A relevant example of reflexive and strictly convex Banach space is $\mathcal{X} = L_q(\mathbb{R} \times \mathbb{S}^{d-1})$ for $q \in (1, \infty)$. Its topological dual is $\mathcal{X}' = L_p(\mathbb{R} \times \mathbb{S}^{d-1})$ with $p = q/(q-1)$ being the conjugate exponent of q. For that particular pair, (89) reduces to the Hölder inequality for hyper-spherical functions. From [15, Chapter 4], the corresponding duality map $\mathrm{J}_q : L_q(\mathbb{R} \times \mathbb{S}^{d-1}) \to L_p(\mathbb{R} \times \mathbb{S}^{d-1})$ is

$$\mathrm{J}_q\{\nu\}(\boldsymbol{z}) = \nu^*(\boldsymbol{z}) = \frac{|\nu(\boldsymbol{z})|^{q-1}}{\|\nu\|_{L_q}^{q-2}} \operatorname{sign}\big(\nu(\boldsymbol{z})\big), \tag{92}$$

which establishes a one-to-one isometric mapping between the spaces L_q and $L_p = (L_q)'$ with the property that $\mathrm{J}_q^{-1} = \mathrm{J}_p$.

Proposition 4 (Banach isometries) *Let $(\mathcal{X}, \mathcal{X}')$ be a dual pair of reflexive and strictly convex Banach spaces with corresponding duality map $\mathrm{J}_{\mathcal{X}} : \mathcal{X} \to \mathcal{X}'$. We consider two generic types of linear isometries.*

1) One-to-one map: Let $\mathrm{T} : \mathcal{X} \to \mathcal{Y} = \mathrm{T}(\mathcal{X})$ be an injective operator whose inverse is denoted by T^{-1} with $\mathrm{T}^{-1}\mathrm{T} = \mathrm{Id}$ on $\mathcal{X}$. Then, $\mathcal{Y} = \mathrm{T}(\mathcal{X}) = \{y = \mathrm{T}\{x\} : x \in \mathcal{X}\}$ equipped with the norm $\|y\|_{\mathcal{Y}} \triangleq \|\mathrm{T}^{-1}\{y\}\|_{\mathcal{X}}$ is a Banach space. Its continuous dual is the Banach space $\mathcal{Y}' = \mathrm{T}^{-1*}(\mathcal{X}')$ with $\|y'\|_{\mathcal{Y}'} = \|\mathrm{T}^*\{y'\}\|_{\mathcal{X}'}$, while the corresponding duality map is $\mathrm{J}_{\mathcal{Y}} = (\mathrm{T}^*)^{-1} \circ \mathrm{J}_{\mathcal{X}} \circ \mathrm{T}^{-1} : \mathcal{Y} \to \mathcal{Y}'$.

2) Projection: Let $\mathrm{P} : \mathcal{X} \to \mathcal{U} \xhookrightarrow{\text{iso.}} \mathcal{X}$ be a continuous projection operator on $\mathcal{X}$ with $\|\mathrm{P}\| = 1$. Then, $\big(\mathrm{P}(\mathcal{X}), \mathrm{P}^*(\mathcal{X}')\big) = \big(\mathcal{U}, \mathcal{U}'\big)$ is a dual pair of Banach subspaces with corresponding duality map $\mathrm{J}_{\mathcal{U}} = \mathrm{P}^* \circ \mathrm{J}_{\mathcal{X}} \circ \mathrm{P} : \mathcal{U} \to \mathcal{U}'$.

Proof We first recall that the dual of a reflexive and strictly convex Banach space is reflexive (by definition) and strictly convex as well.
1) Injective operator: For the first property, we refer to [58, Proposition 1]. The key observation is that the operators $\mathrm{T} : \mathcal{X} \to \mathcal{Y}$ and $\mathrm{T}^{-1} : \mathcal{Y} \to \mathcal{X}$, as well as their adjoints, are isometries with $(\mathrm{T}^*)^{-1} = \mathrm{T}^{-1*}$. The argument then primarily rests upon the duality inequality

$$\begin{aligned}\langle y', y\rangle_{\mathcal{Y}'\times\mathcal{Y}} = \langle y', \mathrm{T}\mathrm{T}^{-1}\{y\}\rangle_{\mathcal{Y}'\times\mathcal{Y}} &= \langle \mathrm{T}^*\{y'\}, \mathrm{T}^{-1}\{y\}\rangle_{\mathcal{X}'\times\mathcal{X}} \\ &\leq \|\mathrm{T}^*\{y'\}\|_{\mathcal{X}'}\ \|\mathrm{T}^{-1}\{y\}\|_{\mathcal{X}},\end{aligned} \tag{93}$$

which is sharp if and only if $x = \mathrm{T}^{-1}\{y\}$ and $x' = \mathrm{T}^*\{y'\}$ (resp., y' and y) are Banach conjugates, so that $x' = \mathrm{J}_{\mathcal{X}}\{x\}$.
2) Projection operator: The first part is obtained by using a standard argument with (projected) Cauchy sequences. For the second part, let $u \in \mathcal{U} \xhookrightarrow{\text{iso.}} \mathcal{X}$ with Banach conjugate $u^* \in \mathcal{X}'$. Then,

$$\begin{aligned}\|u^*\|_{\mathcal{X}'}\|u\|_{\mathcal{X}} = \langle u^*, u\rangle_{\mathcal{X}'\times\mathcal{X}} = \langle u^*, \mathrm{P}^2 u\rangle_{\mathcal{X}'\times\mathcal{X}} &= \langle \mathrm{P}^* u^*, \mathrm{Pu}\rangle_{\mathcal{U}'\times\mathcal{U}} \\ &\leq \|\mathrm{P}^*\|\|u^*\|_{\mathcal{U}'}\|\mathrm{P}\|\|u\|_{\mathcal{U}},\end{aligned} \tag{94}$$

from which we deduce that $\langle \mathrm{P}^* u^*, \mathrm{Pu}\rangle_{\mathcal{U}'\times\mathcal{U}} = \|u^*\|_{\mathcal{U}'}\|u\|_{\mathcal{U}'}$. We conclude that $u = \mathrm{P}u$ and $u^* = \mathrm{P}^* u^* = \mathrm{P}^* \circ \mathrm{J}_{\mathcal{X}}\{\mathrm{P}u\}$ are (unique) Banach conjugates of each other. □

Acknowlegdments

The research was supported by the European Commission under Grant ERC-2020-AdG FunLearn-101020573. The author is thankful to Sebastian Neumayer for very helpful discussions.

References

1. Alvarez, M.A., Rosasco, L., Lawrence, N.D.: Kernels for vector-valued functions: A review. Foundations and Trends in Machine Learning **4**(3), 195–266 (2012)
2. Aronszajn, N.: Theory of reproducing kernels. Transactions of the American Mathematical Society **68**(3), 337–404 (1950)
3. Aziznejad, S., Unser, M.: Multikernel regression with sparsity constraint. SIAM Journal on Mathematics of Data Science **3**(1), 201–224 (2021). DOI 10.1137/20m1318882
4. Bach, F.: Breaking the curse of dimensionality with convex neural networks. Journal of Machine Learning Research **18**, 1–53 (2017)
5. Barron, A.R.: Universal approximation bounds for superpositions of a sigmoidal function. IEEE Transactions on Information Theory **39**(3), 930–945 (1993). DOI 10.1109/18.256500
6. Bartolucci, F., Vito, E.D., Rosasco, L., Vigogna, S.: Understanding neural networks with reproducing kernel Banach spaces. Applied and Computational Harmonic Analysis **62**, 194–236 (2023)
7. Berlinet, A., Thomas-Agnan, C.: Reproducing Kernel Hilbert Spaces in Probability and Statistics, vol. 3. Kluwer Academic Boston (2004)
8. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer (2006)
9. de Boor, C.: On "best" interpolation. Journal of Approximation Theory **16**(1), 28–42 (1976)
10. de Boor, C., Lynch, R.E.: On splines and their minimum properties. Journal of Mathematics and Mechanics **15**(6), 953–969 (1966)
11. Boyer, C., Chambolle, A., De Castro, Y., Duval, V., De Gournay, F., Weiss, P.: On representer theorems and convex regularization. SIAM Journal of Optimization **29**(2), 1260–1281 (2019)
12. Bredies, K., Carioni, M.: Sparsity of solutions for variational inverse problems with finite-dimensional data. Calculus of Variations and Partial Differential Equations **59**(14), 26 (2020)
13. Buhmann, M.D.: Radial Basis Functions. Cambridge University Press (2003)
14. Chen, Z., Haykin, S.: On different facets of regularization theory. Neural Computation **14**(12), 2791–2846 (2002)
15. Cioranescu, I.: Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems, vol. 62. Springer Science & Business Media (2012)
16. Cybenko, G.: Approximation by superpositions of a sigmoidal function. Mathematics of Control, Signals and Systems **2**(4), 303–314 (1989)
17. Duchon, J.: Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In: W. Schempp, K. Zeller (eds.) Constructive Theory of Functions of Several Variables, pp. 85–100. Springer-Verlag, Berlin (1977)
18. Fisher, S.D., Jerome, J.W.: Spline solutions to L_1 extremal problems in one and several variables. Journal of Approximation Theory **13**(1), 73–83 (1975)
19. Gelfand, I.M., Shilov, G.: Generalized Functions. Vol. 1. Properties and Operations. Academic Press, New York (1964)
20. Gelfand, I.M., Shilov, G.: Generalized Functions. Vol. 5. Integral Geometry and Representation Theory. Academic Press, New York (1966)
21. Grafakos, L.: Classical Fourier Analysis. Springer (2008)
22. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
23. Helgason, S.: Integral Geometry and Radon Transforms. Springer (2011)

24. Hofmann, T., Schölkopf, B., Smola, A.J.: Kernel methods in machine learning. Annals of Statistics **36**(3), 1171–1220 (2008)
25. Hornik, K., Stinchcombe, M., White, H.: Multilayer feedforward networks are universal approximators. Neural Networks **2**(5), 359–366 (1989)
26. Kostadinova, S., Pilipović, S., Saneva, K., Vindas, J.: The ridgelet transform of distributions. Integral Transforms and Special Functions **25**(5), 344–358 (2014)
27. Ludwig, D.: The Radon transform on Euclidean space. Communications on Pure and Applied Mathematics **19**(1), 49–81 (1966). DOI 10.1002/cpa.3160190105
28. Mammen, E., van de Geer, S.: Locally adaptive regression splines. Annals of Statistics **25**(1), 387–413 (1997)
29. Meinguet, J.: Multivariate interpolation at arbitrary points made simple. Zeitschrift fur Angewandte Mathematik und Physik **30**, 292–304 (1979)
30. Mhaskar, H., Micchelli, C.A.: Approximation by superposition of sigmoidal and radial basis functions. Advances in Applied Mathematics **13**(3), 350–373 (1992). DOI 10.1016/0196-8858(92)90016-p
31. Micchelli, C.A.: Interpolation of scattered data: Distance matrices and conditionally positive definite functions. Constructive Approximation **2**(1), 11–22 (1986)
32. Micchelli, C.A., Xu, Y., Zhang, H.: Universal kernels. Journal of Machine Learning Research **7**, 2651–2667 (2006)
33. Natterer, F.: The Mathematics of Computed Tomography. John Willey & Sons Ltd. (1984)
34. Neumayer, S., Unser, M.: Explicit representations for Banach subspaces of Lizorkin distributions. Analysis and Applications (in press). DOI 10.1142/S0219530523500148
35. Ongie, G., Willett, R., Soudry, D., Srebro, N.: A function space view of bounded norm infinite width ReLU nets: The multivariate case. International Conference on Representation Learning (ICLR) (2020)
36. Parhi, R., Nowak, R.D.: The role of neural network activation functions. IEEE Signal Processing Letters **27**, 1779–1783 (2020). DOI 10.1109/LSP.2020.3027517
37. Parhi, R., Nowak, R.D.: Banach space representer theorems for neural networks and ridge splines. Journal of Machine Learning Research **22**(43), 1–40 (2021)
38. Parhi, R., Nowak, R.D.: What kinds of functions do deep neural networks learn? Insights from variational spline theory. SIAM Journal on Mathematics of Data Science **4**(2), 464–489 (2022). DOI 10.1137/21m1418642
39. Pinkus, A.: Approximation theory of the MLP model in neural networks. Acta Numerica **8**, 143–195 (1999). DOI 10.1017/s0962492900002919
40. Poggio, T., Girosi, F.: Networks for approximation and learning. Proceedings of the IEEE **78**(9), 1481–1497 (1990)
41. Poggio, T., Girosi, F.: Regularization algorithms for learning that are equivalent to multilayer networks. Science **247**(4945), 978–982 (1990)
42. Poggio, T., Smale, S.: The mathematics of learning: Dealing with data. Notices of the AMS **50**(5), 537–544 (2003)
43. Ramm, A.G., Katsevich, A.I.: The Radon transform and local tomography. CRC Press (2020)
44. Reed, M., Simon, B.: Methods of Modern Mathematical Physics. Vol. 1: Functional Analysis. Academic Press, San Diego (1980)
45. Rudin, W.: Functional Analysis, 2nd edn. McGraw-Hill, New York (1991). McGraw-Hill Series in Higher Mathematics
46. Samko, S.G.: 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)
47. Samko, S.G., Kilbas, A.A., Marichev, O.I.: Fractional Integrals and Derivatives: Theory and Applications. Gordon and Breach Science Publishers (1993)
48. Savarese, P., Evron, I., Soudry, D., Srebro, N.: How do infinite width bounded norm networks look in function space? In: Conference on Learning Theory. PMLR, pp. 2667–2690 (2019)
49. Schölkopf, B., Herbrich, R., Smola, A.J.: A generalized representer theorem. In: D. Helmbold, B. Williamson (eds.) Computational Learning Theory, pp. 416–426. Springer Berlin Heidelberg (2001)

50. Schölkopf, B., Sung, K.K., Burges, C.J.C., Girosi, F., Niyogi, P., Poggio, T., Vapnik, V.: Comparing support vector machines with Gaussian kernels to radial basis function classifiers. IEEE Transactions on Signal Processing **45**(11), 2758–2765 (1997)
51. Schwartz, L.: Théorie des Distributions. Hermann, Paris (1966)
52. Shawe-Taylor, J., Cristianini, N., et al.: Kernel methods for pattern analysis. Cambridge university press (2004)
53. Sonoda, S., Ishikawa, I., Ikeda, M.: Ridge regression with over-parametrized two-layer networks converge to ridgelet spectrum. pp. 2674–2682. PMLR (2021)
54. Sonoda, S., Murata, N.: Neural network with unbounded activation functions is universal approximator. Applied and Computational Harmonic Analysis **43**(2), 233–268 (2017)
55. Trèves, F.: Topological Vector Spaces, Distributions and Kernels. Dover Publications, New York (2006)
56. Unser, M.: A unifying representer theorem for inverse problems and machine learning. Foundations of Computational Mathematics **21**(4), 941–960 (2021). DOI 10.1007/s10208-020-09472-x
57. Unser, M.: Ridges, neural networks, and the Radon transform. Journal of Machine Learning Research **24**, 1–33 (2023)
58. Unser, M., Aziznejad, S.: Convex optimization in sums of Banach spaces. Applied and Computational Harmonic Analysis **56**, 1–25 (2022). DOI 10.1016/j.acha.2021.07.002
59. Unser, M., Blu, T.: Fractional splines and wavelets. SIAM Review **42**(1), 43–67 (2000)
60. Unser, M., Fageot, J., Ward, J.P.: Splines are universal solutions of linear inverse problems with generalized-TV regularization. SIAM Review **59**(4), 769–793 (2017)
61. Unser, M., Tafti, P.D.: An Introduction to Sparse Stochastic Processes. Cambridge University Press (2014)
62. Wahba, G.: Spline Models for Observational Data. Society for Industrial and Applied Mathematics, Philadelphia (1990)
63. Wendland, H.: Scattered Data Approximations. Cambridge University Press (2005)