Analysis and Applications, Vol. 21, No. 5 (2023) 1223–1250

DOI: 10.1142/S0219530523500148

Explicit representations for Banach subspaces of Lizorkin distributions

Sebastian Neumayer* and Michael Unser†

Biomedical Imaging Group
École Polytechnique Fédérale de Lausanne (EPFL)
Station 17, 1015 Lausanne, Switzerland
**sebastian.neumayer@epfl.ch*
†michael.unser@epfl.ch

Received 17 May 2022
Accepted 13 June 2023
Published 27 July 2023

The Lizorkin space is well suited to the study of operators like fractional Laplacians and the Radon transform. In this paper, we show that the space is unfortunately not complemented in the Schwartz space. In return, we show that it is dense in $C_0(\mathbb{R}^d)$, a property that is shared by the larger Schwartz space and that turns out to be useful for applications. Based on this result, we investigate subspaces of Lizorkin distributions that are Banach spaces and for which a continuous representation operator exists. Then, we introduce a variational framework that involves these spaces and that makes use of the constructed operator. By investigating two particular cases of this framework, we are able to strengthen existing results for fractional splines and 2-layer ReLU networks.

Keywords: Fractional splines; inverse problems; Lizorkin space; Lizorkin distributions; quotient spaces; ReLU networks; variational problems.

Mathematics Subject Classification 2020: 46A11, 46B20, 47A52, 49J27, 54B15, 65D07, 68T05

1. Introduction

This paper pertains to the Lizorkin space $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$, which was first introduced in [20] for the investigation of partial differential equations. It consists of the Schwartz functions $\mathcal{S}(\mathbb{R}^d)$ for which all moments vanish. For detailed expositions on the topic, we refer to [26, 28]. Surprisingly, $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ is still rather large, its closure under the L_∞-norm being the space $C_0(\mathbb{R}^d)$. Another attractive feature of $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ is that many non-invertible operators become invertible if their domain is restricted to $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$, which happen with the Radon transform and with fractional Laplacians. This makes $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ well suited to theoretical analyses. In the present paper, we show that $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ cannot be complemented in $\mathcal{S}(\mathbb{R}^d)$. In particular, continuous projections onto $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ cannot exist. This result is in sharp contrast to the periodic setting, where a projection actually exists.

*Corresponding author

The corresponding dual space $\mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$ of Lizorkin distributions [41] is fairly large and has attracted increased interest over the past years (e.g., shearlet transform [2], ridgelet transform [18], choice of activation functions in neural networks [33]). It is well known that $\mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$ can be identified as the quotient space $\mathcal{S}'(\mathbb{R}^d)/\mathcal{P}(\mathbb{R}^d)$, where $\mathcal{P}(\mathbb{R}^d)$ denotes the space of polynomials. Naturally, this leads to a representation problem if we want to make computations explicit, an issue that has not been addressed so far. Here, our results directly imply that no continuous linear projector for the assignment of representatives can exist. At first glance, this result appears to be quite discouraging as it implies that, in general, it is necessary to work with equivalence classes. Fortunately, this can be circumvented if we consider appropriate Banach subspaces of $\mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$. Indeed, due to the density of $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ in $C_0(\mathbb{R}^d)$ and due to the Riesz theorem, the space of Radon measures is an embedded subspace of $\mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$ for which unique representations exist. Further, we are able to provide a positive answer for more general cases if we restrict ourselves to subspaces that can be equipped with a specific Banach-space structure. In this case, we are able to provide a continuous representation operator for which the point evaluations are weak*-continuous.

The procedure to obtain these subspaces and the representatives is as follows: Given a well understood pair of Banach spaces $(\mathcal{X}, \mathcal{X}')$, we construct Banach spaces $(\mathcal{X}_{\mathrm{T}}, \mathcal{X}'_{\mathrm{T}})$ through a linear homeomorphism $\mathrm{T}\colon \mathcal{S}_1 \subset \mathcal{X} \to \mathcal{S}_2$ defined on some dense subspace of $\mathcal{X}$. The core advantage of our construction is that many properties carry over directly to $(\mathcal{X}_{\mathrm{T}}, \mathcal{X}'_{\mathrm{T}})$; for instance, the set of the extreme points of $\mathcal{X}'_{\mathrm{T}}$ can be specified easily from those of $\mathcal{X}'$. While our construction is abstract, it allows us to make use of the fact that many (differential) operators are homeomorphisms on $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ due to the density of $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ in $\mathcal{X} = C_0(\mathbb{R}^d)$. Unfortunately, the space $\mathcal{X}'_{\mathrm{T}}$ usually still consists of mere equivalence classes. Therefore, as the second step, we formulate conditions under which we can identify the elements of $\mathcal{X}'_{\mathrm{T}}$ using a representation operator. These conditions are fulfilled when the Green's function of the operator T is sufficiently regular. Overall, this framework enables us to design a rich class of interesting new norms for which the related Banach subspaces of Lizorkin distributions have a continuous representation operator. This makes our framework usable for applications.

Within the proposed setting, we study variational problems involving the constructed Banach spaces and the general representer theorems established in [37, 38]. The fact that our abstract formulation involves spaces whose elements are equivalence classes can be circumvented by the application of our representation operator. We investigate two special cases for which the formulations become explicit. First, we revisit fractional splines in arbitrary dimensions, which have been investigated before in [9, 39, 40]. These splines are a generalization of the traditional polynomial splines [31] and preserve most of their properties. Note that (fractional) splines are still fashionable and recently found their way into neural network research [3, 11, 24]. Although this is not included in our discussion, we may also use the model to study polynomial splines. Overall, our approach leads to a unifying setting that includes a straightforward extension to the multivariate case.

As second example, we strengthen the representation results for 2-layer ReLU networks established by Parhi and Nowak [23] and Bartolucci *et al.* [1], which builds up on the univariate case investigated in [30]. The involved norm was also studied from a theoretical point of view in [22]. While proofs in these works rely on a general result on the existence of sparse solutions for variational problems by Bredies and Carioni [4], we are additionally able to identify the solution set as being the weak* closure of certain sparse solutions. Similarly to [1], our construction is related to reproducing-kernel Banach spaces [6, 19, 42] since our representation operator is constructed using a kernel and the point evaluations are continuous. The key ingredient that enables us to strengthen the results of these prior works is that we actually construct a predual space for the optimization domain, which enables us to use our proposed variational framework.

The paper is organized as follows: The necessary preliminaries are provided in Sec. 2. Then, we proceed with a discussion of $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ in Sec. 3 and show that a continuous projection onto $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ cannot exist. This part is complemented with a short discussion of the periodic case. In Sec. 4, we identify subspaces of Lizorkin distributions for which a continuous representation operator exists. Next, we relate these subspaces to several interesting research questions in Sec. 5. As warm-up, we investigate the construction of periodic (fractional) splines in Sec. 5.1. Here, no representation mechanism is necessary as we can use the projector. Then, we introduce in Sec. 5.2 our general variational framework involving the constructed Banach spaces, for which we detail two specific cases: The delicate case of non-periodic (fractional) splines in Sec. 5.3; and a representer theorem for 2-layer ReLU neural networks in Sec. 5.4. Finally, conclusions are drawn in Sec. 6.

2. Mathematical Preliminaries

In this paper, we consider functions $f\colon \mathbb{R}^d \to \mathbb{C}$. To describe their partial derivatives, we use the multi-index $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{N}^d$ with the notational conventions $\mathbf{k}! = \prod_{n=1}^d k_n!$, $|\mathbf{k}| = k_1 + \cdots + k_d$, $\mathbf{x}^{\mathbf{k}} = \prod_{n=1}^d x_n^{k_n}$ for any $\mathbf{x} \in \mathbb{R}^d$ and

$$\partial^{\mathbf{k}} f(\mathbf{x}) = \frac{\partial^{|\mathbf{k}|} f(x_1, \dots, x_d)}{\partial_{x_1}^{k_1} \cdots \partial_{x_d}^{k_d}}. \tag{2.1}$$

This enables us to write the multidimensional Taylor expansion around $\mathbf{x}_0$ of an analytical function $f\colon \mathbb{R}^d \to \mathbb{C}$ in compact form as

$$f(\mathbf{x}) = \sum_{n=0}^{\infty} \sum_{|\mathbf{k}|=n} \frac{\partial^{\mathbf{k}} f(\mathbf{x}_0)}{\mathbf{k}!} (\mathbf{x} - \mathbf{x}_0)^{\mathbf{k}}. \tag{2.2}$$

The Schwartz space [32] of smooth and rapidly decreasing functions $\varphi\colon \mathbb{R}^d \to \mathbb{C}$ equipped with the usual Fréchet–Schwartz topology is denoted by $\mathcal{S}(\mathbb{R}^d)$. This space is an algebra for the multiplication as well as the convolution product. Additionally, it is closed under translation, differentiation and multiplication by polynomials. Its continuous dual is the space of tempered distributions $\mathcal{S}'(\mathbb{R}^d)$. Moreover, $\mathcal{S}'(\mathbb{R}^d)$ (as well as $\mathcal{S}(\mathbb{R}^d)$) is a nuclear Montel space, where a sequence in $\mathcal{S}'(\mathbb{R}^d)$ converges with respect to the strong dual topology if and only if it converges in the weak* topology. Therefore, it does not actually matter which of the two topologies we choose for $\mathcal{S}'(\mathbb{R}^d)$. The Montel property also implies that $\mathcal{S}(\mathbb{R}^d)$ is reflexive, which means that there exists an isomorphism between the topological vector spaces $\mathcal{S}''(\mathbb{R}^d)$ and $\mathcal{S}(\mathbb{R}^d)$. Note that the Lebesgue spaces $L_p(\mathbb{R}^d)$ for $p \in [1, \infty)$ are the completion of the set $\mathcal{S}(\mathbb{R}^d)$ under the L_p-norm $\|\cdot\|_{L_p}$. For $p = \infty$, we have that $\overline{(\mathcal{S}(\mathbb{R}^d), \|\cdot\|_{L_\infty})} = C_0(\mathbb{R}^d)$, namely, the space of continuous functions that vanish at infinity. The 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 norm

$$\|f\|_{\mathcal{M}} = \sup_{\varphi \in \mathcal{S}(\mathbb{R}^d) : \|\varphi\|_{L_\infty} \leq 1} \langle f, \varphi \rangle. \tag{2.3}$$

The latter is an isometrically-embedded superset of $L_1(\mathbb{R}^d)$, which implies that $\|f\|_{L_1} = \|f\|_{\mathcal{M}}$ for all $f \in L_1(\mathbb{R}^d)$. We also need the weighted Lebesgue space $L_{\infty,\alpha}(\mathbb{R}^d)$, $\alpha \geq 0$, defined via the weighted norm

$$\|f\|_{\infty,\alpha} \coloneqq \operatorname*{ess\,sup}_{\mathbf{x} \in \mathbb{R}^d} |f(\mathbf{x})|(1 + \|\mathbf{x}\|_2)^{-\alpha}, \tag{2.4}$$

which consists of functions that grow with order at most α.

The Fourier transform $\mathcal{F} \colon L_1(\mathbb{R}^d) \to C_0(\mathbb{R}^d)$ of a function $\varphi \in L_1(\mathbb{R}^d)$ is defined as

$$\widehat{\varphi}(\boldsymbol{\omega}) \coloneqq \mathcal{F}\{\varphi\}(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi(\mathbf{x}) \mathrm{e}^{-\mathrm{j}\langle \boldsymbol{\omega}, \mathbf{x} \rangle} \mathrm{d}\mathbf{x}. \tag{2.5}$$

As the Fourier transform $\mathcal{F} \colon \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$ is an isomorphism, it can be extended by duality to $\mathcal{S}'(\mathbb{R}^d)$. Specifically, $\widehat{f} \in \mathcal{S}'(\mathbb{R}^d)$ 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 $\varphi \in \mathcal{S}(\mathbb{R}^d)$. Finally, we note that the analytic Schwartz functions form a dense subset of $\mathcal{S}(\mathbb{R}^d)$, which can be seen as follows. As the smooth and compactly supported functions $\mathcal{D}(\mathbb{R}^d)$ are dense in $\mathcal{S}(\mathbb{R}^d)$, we also get that $\mathcal{F}(\mathcal{D}(\mathbb{R}^d))$ is dense in $\mathcal{S}(\mathbb{R}^d)$. Due to the Paley–Wiener theorem, the Fourier transform of any $f \in \mathcal{D}(\mathbb{R}^d)$ is analytic and also entire. Hence, these functions are dense.

The simplest way to specify fractional derivatives or integrals is to describe their action in the Fourier domain. Let us start with $d = 1$. The one-dimensional fractional derivative $\mathrm{D}^\alpha \colon \mathcal{S}(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$ of order $\alpha \geq 0$ is defined as

$$\mathrm{D}^\alpha\{\varphi\}(t) = \mathcal{F}^{-1}\{(\mathrm{j}\cdot)^\alpha \widehat{\varphi}\}(t). \tag{2.6}$$

For $\alpha = n \in \mathbb{N}$, $\mathrm{D}^n = \frac{\mathrm{d}^n}{\mathrm{d}t^n}$ coincides with the classical nth order derivative. Definition (2.6) is also valid for negative orders, in which case it yields a fractional integral [39]. In fact, the impulse response of $\mathrm{D}^{-\alpha}$ is the Green's function of D^{α}, which is given by

$$\rho_\alpha(t) = \mathcal{F}^{-1}\left\{\frac{1}{(\mathrm{j}\cdot)^{-\alpha}}\right\}(t) = \begin{cases} \dfrac{t_+^{\alpha-1}}{\Gamma(\alpha)}, & \alpha-1 \in \mathbb{R}^+\backslash\mathbb{N}, \\ \dfrac{\mathrm{sgn}(t)}{2}\dfrac{t^n}{n!}, & \alpha-1 = n \in \mathbb{N}. \end{cases} \tag{2.7}$$

Likewise, the fractional Laplacian $(-\Delta)^{\alpha/2}$ of order $\alpha \in (1,\infty)$ is the linear-shift-invariant operator (LSI) whose frequency response is $\|\boldsymbol{\omega}\|^{\alpha}$. Its inverse is the fractional integrator $(-\Delta)^{-\alpha/2}$, which corresponds to a frequency-domain multiplication by $\|\boldsymbol{\omega}\|^{-\alpha}$. Fractional derivatives and Laplacians are part of the same family of operators (isotropic LSI and scale-invariant) with their distributional impulse response for $\alpha > d$ being given by

$$k_{\alpha,d}(\mathbf{x}) = \mathcal{F}^{-1}\left\{\frac{1}{\|\cdot\|^{\alpha}}\right\}(\mathbf{x}) = \begin{cases} (-\Delta)^{-n}\{\delta\}, & \alpha/2 = n \in \mathbb{N} \\ B_{n,d}\|\mathbf{x}\|^{2n}\log(\|\mathbf{x}\|), & \alpha - d = 2n \in 2\mathbb{N} \\ A_{\alpha,d}\|\mathbf{x}\|^{\alpha-d}, & \alpha - d \notin 2\mathbb{N} \end{cases} \tag{2.8}$$

with constants $A_{d,\alpha} = \frac{\Gamma((d-\alpha)/2)}{2^{\alpha}\pi^{d/2}\Gamma(\alpha/2)}$ and $B_{d,n} = \frac{(-1)^{1+n}}{2^{2n+d-1}\pi^{d/2}\Gamma(n+d/2)n!}$. For a more detailed exposition on the topic, we refer to [13, 29, 34].

3. Lizorkin Spaces

The Lizorkin space $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ is the closed subspace of $\mathcal{S}(\mathbb{R}^d)$ that consists of the functions whose moments of any order $\mathbf{k}$ are zero, so that

$$\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d) = \left\{\varphi \in \mathcal{S}(\mathbb{R}^d) : \int_{\mathbb{R}^d} \mathbf{x}^{\mathbf{k}}\varphi(\mathbf{x})\mathrm{d}\mathbf{x} = 0,\ \forall \mathbf{k} \in \mathbb{N}^d\right\}. \tag{3.1}$$

A nice overview with properties of $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ is given in [36]. Equivalently, we can describe these functions in the Fourier domain through

$$\widehat{\mathcal{S}}_{\mathrm{Liz}}(\mathbb{R}^d) = \mathcal{F}(\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)) = \{\psi \in \mathcal{S}(\mathbb{R}^d) : \partial^{\mathbf{k}}\psi(\mathbf{0}) = 0\ \forall \mathbf{k} \in \mathbb{N}^d\}. \tag{3.2}$$

Although closed subspaces of reflexive topological vector spaces are in general not reflexive, this property holds for Fréchet spaces. Hence, the spaces $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ and $\widehat{\mathcal{S}}_{\mathrm{Liz}}(\mathbb{R}^d)$ are reflexive. Further, we have for all $\varphi \in \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$, $\mathbf{x}_0 \in \mathbb{R}^d$ and $a \in \mathbb{R}$ that $\varphi(\cdot - \mathbf{x}_0) \in \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ and $\varphi(\cdot/a) \in \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$. Finally, we note that $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d) \cap \mathcal{D}(\mathbb{R}^d) = \{0\}$. Indeed, if $\varphi \in \mathcal{D}(\mathbb{R}^d)$, then $\widehat{\varphi}$ is entire and hence equal to its Maclaurin expansion. But if $\varphi \in \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$, then the Taylor series of $\widehat{\varphi}$ is 0.

We are going to show that $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ cannot be complemented in $\mathcal{S}(\mathbb{R}^d)$; in other words, a continuous projector $\mathrm{P}_{\mathrm{Liz}} \colon \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ cannot exist. Before we prove this negative result, we first discuss the easier case of periodic Lizorkin functions, for which a continuous projection actually exists.

3.1. *Periodic Lizorkin spaces*

The functions of interest are T-periodic and typically specified only over their main period $\mathbb{T} = [0, T]$. The corresponding space of test functions is $\mathcal{S}(\mathbb{T}) = C^\infty_{\mathrm{perio}}(\mathbb{T})$, which is in one-to-one correspondence with the Fréchet space of rapidly decaying sequences[a] $\mathcal{S}(\mathbb{Z})$ via the Fourier homeomorphism [10, 35]. More precisely, there are Fourier coefficients $\widehat{\varphi}[\cdot] \in \mathcal{S}(\mathbb{Z})$ such that

$$\varphi(t) = \sum_{n\in\mathbb{Z}} \widehat{\varphi}[n]\mathrm{e}^{\mathrm{j}n\omega_0 t} \in \mathcal{S}(\mathbb{T}) \tag{3.3}$$

with $\omega_0 = \frac{2\pi}{T}$. This expansion of φ is unique and $\widehat{\varphi}[n] = \langle \varphi, \mathrm{e}^{-\mathrm{j}n\omega_0\cdot}\rangle_{\mathbb{T}}$, where $\langle f, g\rangle_{\mathbb{T}} = \frac{1}{T}\int_{\mathbb{T}} f(t)g(t)\mathrm{d}t$. The continuous dual of $\mathcal{S}(\mathbb{T})$ is the space of periodic distributions $\mathcal{S}'(\mathbb{T}) = \mathcal{S}'_{\mathrm{perio}}(\mathbb{R})$, which is itself homeomorphic to the space $\mathcal{S}'(\mathbb{Z})$ of slowly growing sequences. Indeed, it holds $f \in \mathcal{S}'(\mathbb{T}) \Leftrightarrow \widehat{f}[\cdot] \in \mathcal{S}'(\mathbb{Z})$, where $\widehat{f}[n]$ denotes the nth Fourier coefficient of f.

To ensure invertibility of the continuous fractional-derivative operator $\mathrm{D}^\alpha\colon \mathcal{S}(\mathbb{T}) \to \mathcal{S}(\mathbb{T})$ given by

$$\mathrm{D}^\alpha\{\varphi\}(t) = \sum_{n\in\mathbb{Z}} (\mathrm{j}\omega_0 n)^\alpha \widehat{\varphi}[n]\mathrm{e}^{\mathrm{j}n\omega_0 t}, \tag{3.4}$$

we restrict ourselves to the subspace

$$\mathcal{S}_0(\mathbb{T}) = \{\varphi \in \mathcal{S}(\mathbb{T}) : \langle 1, \varphi\rangle_{\mathbb{T}} = 0\}, \tag{3.5}$$

which inherits the nuclear topology from $\mathcal{S}(\mathbb{T})$. While (3.5) imposes a restriction only on the mean value of φ, the resulting space $\mathcal{S}_0(\mathbb{T})$ is the proper periodic counterpart of the $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R})$, since the only periodic polynomials are constants. The periodic setting is simple, in that $\mathcal{S}_0(\mathbb{T})$ is 1-complemented in $\mathcal{S}(\mathbb{T})$ with $\mathcal{S}(\mathbb{T}) = \mathcal{S}_0(\mathbb{T}) \oplus \mathcal{P}_0$ and

$$\mathcal{P}_0 = \{b_0 \cdot 1 : b_0 \in \mathbb{R}\} \subset \mathcal{S}(\mathbb{T}). \tag{3.6}$$

Correspondingly, we introduce the continuous projection $\mathrm{P}_0\colon \mathcal{S}(\mathbb{T}) \to \mathcal{S}_0(\mathbb{T})$ with

$$\mathrm{P}_0\{\phi\} = \varphi - \langle 1, \varphi\rangle_{\mathbb{T}} 1. \tag{3.7}$$

As $\mathcal{S}_0(\mathbb{T}) = \mathrm{P}_0(\mathcal{S}(\mathbb{T}))$, its dual is $\mathcal{S}'_0(\mathbb{T}) = \mathrm{P}^*_0(\mathcal{S}'(\mathbb{T}))$ with $\mathcal{S}'(\mathbb{T}) = \mathcal{S}'_0(\mathbb{T}) \oplus \mathcal{P}_0$. Here, we can identify $\mathcal{P}'_0 = \mathcal{P}_0$ because the space is spanned by $1 \in \mathcal{S}(\mathbb{T}) \subset \mathcal{S}'(\mathbb{T})$ with $\langle 1, 1\rangle_{\mathbb{T}} = 1$. The latter property also implies that $\mathrm{P}^*_0 = \mathrm{P}_0$, which makes the projection (3.7) also applicable to periodic distributions.

As the space $\mathcal{P}_0$ of constant polynomials is indeed the null space of $\mathrm{D}^\alpha\colon \mathcal{S}(\mathbb{T}) \to \mathcal{S}(\mathbb{T})$ with $\alpha > 0$, we can restrict D^α to a homeomorphism $\mathrm{D}^\alpha\colon \mathcal{S}_0(\mathbb{T}) \to \mathcal{S}_0(\mathbb{T})$ for any $\alpha \in \mathbb{R}$. By duality, the same holds true on $\mathcal{S}'_0(\mathbb{T})$ with the Fourier-domain

[a]This space is denoted by "s" in [14]. It is the discrete analog of the Schwartz space $\mathcal{S}(\mathbb{R})$.

definition (3.4) of D^α being applicable to periodic distributions as well. In particular, the fractional integrator $\mathrm{D}^{-\alpha}\colon \mathcal{S}'_0(\mathbb{T}) \to \mathcal{S}'_0(\mathbb{T})$ of order $\alpha \geq 0$ is given by

$$\mathrm{D}^{-\alpha}\{f\}(t) = \sum_{n\in\mathbb{Z}\setminus\{0\}} \frac{1}{(\mathrm{j}n\omega_0)^\alpha} \widehat{f}[n]\mathrm{e}^{\mathrm{j}n\omega_0 t}. \tag{3.8}$$

3.2. *Nonexistence of a continuous projector* $\mathbf{P}_{\mathbf{Liz}}\colon \boldsymbol{\mathcal{S}}(\mathbb{R}^{\boldsymbol{d}}) \to \boldsymbol{\mathcal{S}}_{\mathbf{Liz}}(\mathbb{R}^{\boldsymbol{d}})$

To prove the nonexistence of a continuous linear projection onto the Lizorkin space $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$, we first show that the closed set $\mathcal{P}(\mathbb{R}^d)$ is not complemented in $\mathcal{S}'(\mathbb{R}^d)$.

Theorem 3.1. *There exists no topological complement of* $\mathcal{P}(\mathbb{R}^d)$ *in* $\mathcal{S}'(\mathbb{R}^d)$.

Proof. Assume there is a complement. In other words, assume that a continuous projector $\mathrm{P}\colon \mathcal{S}'(\mathbb{R}^d) \to \mathcal{P}(\mathbb{R}^d) \subset \mathcal{S}'(\mathbb{R}^d)$ exists. We consider $\widehat{\delta}_{\mathbf{x}_0} \in \mathcal{S}'(\mathbb{R}^d)$, $\mathbf{x}_0 \in \mathbb{R}^d$, and set $p_{\mathbf{x}_0} \coloneqq \mathrm{P}\{\widehat{\delta}_{\mathbf{x}_0}\} \in \mathcal{P}(\mathbb{R}^d)$. Using relations such as

$$\lim_{t\to 0} \frac{\delta_{te_k} - \delta}{t} = -\nabla_{e_k}\delta \tag{3.9}$$

and similar ones for higher-order derivatives, we observe that $\mathcal{P}(\mathbb{R}^d) \subset \overline{\mathrm{span}\{\widehat{\delta}_{\mathbf{x}_0}\}_{\|\mathbf{x}_0\|\leq 1}}$, where all equalities are in the sense of distributions. Since P is a continuous projection onto $\mathcal{P}(\mathbb{R}^d)$, this implies that

$$\mathcal{P}(\mathbb{R}^d) \subset \overline{\mathrm{span}\{p_{\mathbf{x}_0}\}_{\|\mathbf{x}_0\|\leq 1}}. \tag{3.10}$$

In the sequel, we show that the polynomials $p_{\mathbf{x}_0}$ with $\|\mathbf{x}_0\|_2 \leq 1$ have a common maximum degree m, which results in the contradiction that $\mathcal{P}(\mathbb{R}^d) \subset \mathcal{P}_m(\mathbb{R}^d)$.

If no common maximum exists, then there is a sequence $\{\mathbf{h}_n\}_{n\in\mathbb{N}} \in \mathbb{R}^d$ with $\|\mathbf{h}_n\|_2 \leq 1$ such that $\{p_{\mathbf{h}_n}\}_{n\in\mathbb{N}}$ is a sequence of polynomials with unbounded degree. By passing to a subsequence, we can assume that $\mathbf{h}_n \to \mathbf{h}$ for some $\mathbf{h} \in \mathbb{R}^d$ with $\|\mathbf{h}\| \leq 1$. Due to the continuity of P and $\mathcal{F}$, we also have that $\widehat{p}_{\mathbf{h}_n} \to \widehat{p}_{\mathbf{h}}$ in the sense of distributions. Setting $p_{\mathbf{h}_n} = \sum_{j=0}^{m_n} a_{j,n}x^j$ with $m_n \to \infty$ and $a_{m_n,n} \neq 0$, this can be written as

$$\widehat{p}_{\mathbf{h}_n} = \sum_{j=0}^{m_n} (-2\pi\mathrm{j})^j a_{j,n} \frac{\partial^j}{\partial\xi^j}\delta_0 \to \widehat{p}_{\mathbf{h}}. \tag{3.11}$$

Dropping again to a subsequence, we assume that m_n is monotonically increasing. Using Borel's theorem, we then pick $\varphi \in \mathcal{S}(\mathbb{R}^d)$ with $\frac{\partial^{m_n}}{\partial\xi^{m_n}}\varphi(0) = (a_{m_n,n})^{-1}C_n$, where C_n is chosen such that

$$\left|\sum_{j=0}^{m_n} (-2\pi\mathrm{j})^j a_{j,n} \frac{\partial^j}{\partial\xi^j}\varphi(0)\right| \geq n. \tag{3.12}$$

Hence, $\widehat{p}_{\mathbf{h}_n}(\varphi) \to \infty$, which contradicts that $\widehat{p}_{\mathbf{h}_n} \to \widehat{p}_{\mathbf{h}} \in \mathcal{S}'(\mathbb{R}^d)$. Consequently, all $p_{\mathbf{h}}$ with $\|\mathbf{h}\| \leq 1$ have a common maximum degree m. □

Remark 3.1. Theorem 3.1 implies that there is no continuous linear projection $\mathrm{P}\colon \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ with $\ker \mathrm{P} = \mathcal{P}(\mathbb{R}^d)$. Otherwise, we would have that $(\mathrm{Id} - \mathrm{P})$ is a continuous projector onto $\mathcal{P}(\mathbb{R}^d)$. In particular, representatives of Lizorkin distributions cannot be assigned in a continuous linear way.

Now, the desired nonexistence result follows immediately.

Corollary 3.1. *There exists no topological complement of* $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ *in* $\mathcal{S}(\mathbb{R}^d)$.

Proof. On the contrary, let us assume that a continuous linear projection $\mathrm{P}_{\mathrm{Liz}}\colon \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ exists. Then, the adjoint map $\mathrm{P}^*_{\mathrm{Liz}}\colon \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ is a projection as well. Due to the fact that

$$\langle \mathrm{P}^*_{\mathrm{Liz}}\{f\}, \varphi\rangle = \langle f, \mathrm{P}_{\mathrm{Liz}}\{\varphi\}\rangle \tag{3.13}$$

for all $f \in \mathcal{S}'(\mathbb{R}^d)$ and $\varphi \in \mathcal{S}(\mathbb{R}^d)$, its null space is given by $\ker \mathrm{P}^*_{\mathrm{Liz}} = \mathcal{P}(\mathbb{R}^d)$. Hence, $(\mathrm{Id} - \mathrm{P}^*_{\mathrm{Liz}})$ would be a projection onto $\mathcal{P}(\mathbb{R}^d)$, which contradicts Theorem 3.1. □

3.3. *Closure of the Lizorkin space*

Despite the negative finding of Sec. 3.2, we are nevertheless able to provide a result that is useful for applications.

Theorem 3.2. *It holds that* $\overline{(\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d), \|\cdot\|_\infty)} = C_0(\mathbb{R}^d)$.

We note that the result was already mentioned in [27], but without a proof.

Proof. Using the function $\tilde{\varphi}_0\colon \mathbb{R}^d \to \mathbb{R}$ with $\tilde{\varphi}_0(\mathbf{x}) = \exp(-1/(1-(2\|\mathbf{x}\|)^2))/n_d$ (n_d is the normalizing constant) for $\|\mathbf{x}\| < 1/2$ and zero else, we define $\varphi_0\colon \mathbb{R}^d \to [0,1]$ via $\varphi(\mathbf{x}) = (\chi_{B_1(0)} * \tilde{\varphi}_0)(2\mathbf{x})$ with $\chi_{B_1(0)}$ being the characteristic function of the unit ball. This function is smooth, symmetric, and satisfies that $\varphi(\mathbf{x}) = 1$ for $|\mathbf{x}| \leq 1/4$ and $\varphi(\mathbf{x}) = 0$ for $|\mathbf{x}| \geq 3/4$. Based on this function, we define

$$\phi_{\mathbf{n}} = \frac{\mathbf{x}^{\mathbf{n}}}{\mathbf{n}!}\varphi_0 \in \mathcal{S}(\mathbb{R}^d), \quad \mathbf{n} \in \mathbb{N}^d \tag{3.14}$$

and set

$$E_2 = \overline{\mathrm{span}\{\phi_{\mathbf{n}} : \mathbf{n} \in \mathbb{N}^d\}} \subset \mathcal{S}(\mathbb{R}^d). \tag{3.15}$$

Next, we observe that $(\widehat{\phi}_{\mathbf{n}} - t^{-|\mathbf{n}|-d}\widehat{\phi}_{\mathbf{n}}(\cdot/t)) \to \widehat{\phi}_{\mathbf{n}} \in C_0(\mathbb{R}^d)$ as $t \to \infty$, where the sequences have all moments equal to zero. Hence, we have that

$$\widehat{E}_2 \in \overline{(\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d), \|\cdot\|_\infty)}. \tag{3.16}$$

To conclude the argument, we show that $\widehat{\mathcal{S}}_{\mathrm{Liz}}(\mathbb{R}^d) + E_2$ contains the entire Schwartz functions, so that its closure under the Schwartz topology is already the complete space $\mathcal{S}(\mathbb{R}^d)$. Then, the same holds true for $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d) + \widehat{E}_2$ and consequently, we get that

$$\overline{(\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d), \|\cdot\|_\infty)} = \overline{(\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d) + \widehat{E}_2, \|\cdot\|_\infty)} = \overline{(\mathcal{S}(\mathbb{R}^d), \|\cdot\|_\infty)} = C_0(\mathbb{R}^d). \tag{3.17}$$

The Taylor series of any entire function f converges absolutely for any $\mathbf{x} \in \mathbb{R}^d$. It then holds that

$$g = f\varphi_0 = \sum_{\mathbf{n}\in\mathbb{N}^d} \partial^{\mathbf{n}} f(\mathbf{0})\phi_{\mathbf{n}} \in \mathcal{S}(\mathbb{R}^d). \tag{3.18}$$

Hence, we get that $(f - g) \in \widehat{\mathcal{S}}_{\mathrm{Liz}}(\mathbb{R}^d)$. It remains to show that $g \in E_2$ or, equivalently, that

$$g_j = \sum_{\mathbf{n}\in\mathbb{N}^d, |\mathbf{n}|<j} \partial^{\mathbf{n}} f(\mathbf{0})\phi_{\mathbf{n}} \to g \tag{3.19}$$

in the Schwartz topology. For any $\alpha, \mathbf{k} \in \mathbb{N}^d$, it holds that

$$\begin{aligned}
\|\mathbf{x}^\alpha \partial^{\mathbf{k}}(g - g_j)\|_\infty &\le \sup_{|\mathbf{x}|\le 3/4} |\mathbf{x}^\alpha| \sum_{\mathbf{n}\in\mathbb{N}^d, |\mathbf{n}|\ge j} \left| \frac{\partial^{\mathbf{n}} f(\mathbf{0})}{\mathbf{n}!} \partial^{\mathbf{k}}(\mathbf{x}^{\mathbf{n}}\varphi_0(\mathbf{x})) \right| \\
&\le \sup_{|\mathbf{x}|\le 3/4} \sum_{\mathbf{n}\in\mathbb{N}^d, |\mathbf{n}|\ge j} |\partial^{\mathbf{n}} f(\mathbf{0})| \sum_{\mathbf{k}_1\le \mathbf{k},\mathbf{n}} \binom{\mathbf{k}}{\mathbf{k}_1} \frac{|\mathbf{x}^{\mathbf{n}-\mathbf{k}_1}|}{(\mathbf{n}-\mathbf{k}_1)!} |\partial^{\mathbf{k}-\mathbf{k}_1}\varphi_0(\mathbf{x})| \\
&\le C \sum_{\mathbf{n}\in\mathbb{N}^d, |\mathbf{n}|\ge j} |\partial^{\mathbf{n}} f(\mathbf{0})| \sum_{\mathbf{k}_1\le \mathbf{k},\mathbf{n}} \binom{\mathbf{k}}{\mathbf{k}_1} \frac{1}{(\mathbf{n}-\mathbf{k}_1)!} \\
&\le C \sum_{\mathbf{n}\in\mathbb{N}^d, |\mathbf{n}|\ge j} |\partial^{\mathbf{n}} f(\mathbf{0})| \frac{\mathbf{n}^{\mathbf{k}}}{\mathbf{n}!}.
\end{aligned} \tag{3.20}$$

The last expression converges to zero as $j \to \infty$ if

$$1 < \limsup_{j\to\infty} \left(\sum_{|\mathbf{n}|=j} \frac{|\partial^{\mathbf{n}} f(\mathbf{0})|}{\mathbf{n}!} j^{|\mathbf{k}|} \right)^{-\frac{1}{j}}. \tag{3.21}$$

However, it holds that $\limsup_{j\to\infty} j^{\frac{1}{j}} = 1$ and hence, the claim follows since the Taylor expansion converges absolutely for any $\mathbf{x} \in \mathbb{R}^d$. □

By duality, Theorem 3.2 implies that the Radon measures $\mathcal{M}(\mathbb{R}^d)$ are continuously embedded into the space $\mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$ of Lizorkin distributions.

Remark 3.2. In [27], it was shown that the same results hold for $L_p(\mathbb{R}^d)$ with $1 \le p < \infty$.

4. Banach Subspaces of Lizorkin Distributions

In contrast to the periodic case, the space $\mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$ is an abstract space of equivalence classes, which means that the assignment of representatives for computational purposes is difficult. Therefore, we want to restrict our attention to subspaces with more structure. As our proposed framework is also applicable for other spaces, we outline it in full generality and explicitly provide the specifications for $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ as discussion.

Let $\mathcal{S}_1$, $\mathcal{S}_2$ be two topological vector spaces with a linear homeomorphism $\mathrm{T}\colon \mathcal{S}_1 \to \mathcal{S}_2$ and let $\mathrm{T}^*\colon \mathcal{S}'_2 \to \mathcal{S}'_1$ be defined via duality. The simplest choice for the construction of Banach subspaces of $\mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$ is $\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$, but variations of this setting are clearly possible. In the Lizorkin setting, a possible choice of operator is the fractional Laplacian $\mathrm{T} = (-\Delta)^\alpha$, which is discussed as first example in Sec. 4.1. Now, let $(\mathcal{X}, \mathcal{X}')$ be a dual pair of Banach spaces whose norm $\|\cdot\|_{\mathcal{X}}$ is continuous with respect to the topology of $\mathcal{S}_1$ such that $\overline{(\mathcal{S}_1, \|\cdot\|_{\mathcal{X}})} = \mathcal{X}$. For $\mathcal{S}_1 = \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$, we have seen in Sec. 3.3 that $\mathcal{X} = L_p(\mathbb{R}^d)$, $p \in [1, \infty)$ and $\mathcal{X} = C_0(\mathbb{R}^d)$ are admissible choices, as their norms are indeed compatible with the Schwartz topology. The density enables us to write that

$$\|f\|_{\mathcal{X}'} = \sup_{\varphi \in \mathcal{X}: \|\varphi\|_{\mathcal{X}} \le 1} \langle f, \varphi\rangle = \sup_{\varphi \in \mathcal{S}_1: \|\varphi\|_{\mathcal{X}} \le 1} \langle f, \varphi\rangle \tag{4.1}$$

for any $f \in \mathcal{X}'$. Given any $f \in \mathcal{S}'_1$ for which (4.1) is finite, the bounded linear transformation (BLT) theorem implies that there exists a unique continuous extension to some element in $\mathcal{X}'$. Conversely, any $f \in \mathcal{X}'$ defines a unique element in $\mathcal{S}'_1$ via restriction due to the compatibility of the norm and the topology.

In this setting, we define the abstract space

$$\mathcal{X}'_{\mathrm{T}} = \bigl\{f \in \mathcal{S}'_2 : \|\mathrm{T}^*\{f\}\|_{\mathcal{X}'} < \infty\bigr\} = \bigl\{\mathrm{T}^{-*}\{g\} \in \mathcal{S}'_2 : g \in \mathcal{X}'\bigr\}, \tag{4.2}$$

which is a Banach space if equipped with the norm $\|\cdot\|_{\mathcal{X}'_{\mathrm{T}}} := \|\mathrm{T}^*\{\cdot\}\|_{\mathcal{X}'}$. In particular, by choosing $\mathcal{S}_2 = \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$, we can construct a subspace of $\mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$ and equip it with a Banach-space structure. The norm of $\mathcal{X}'_{\mathrm{T}}$ can be rewritten in dual form as

$$\begin{aligned}\|f\|_{\mathcal{X}'_{\mathrm{T}}} &= \sup_{\varphi \in \mathcal{S}_1: \|\varphi\|_{\mathcal{X}} \le 1} \langle \mathrm{T}^*\{f\}, \varphi\rangle = \sup_{\varphi \in \mathcal{S}_2: \|\mathrm{T}^{-1}\{\varphi\}\|_{\mathcal{X}} \le 1} \langle \mathrm{T}^*\{f\}, \mathrm{T}^{-1}\{\varphi\}\rangle \\ &= \sup_{\varphi \in \mathcal{S}_2: \|\mathrm{T}^{-1}\{\varphi\}\|_{\mathcal{X}} \le 1} \langle f, \mathrm{T}\mathrm{T}^{-1}\{\varphi\}\rangle = \sup_{\varphi \in \mathcal{S}_2: \|\mathrm{T}^{-1}\{\varphi\}\|_{\mathcal{X}} \le 1} \langle f, \varphi\rangle. \end{aligned} \tag{4.3}$$

Consequently, from the BLT theorem again, any $f \in \mathcal{X}'_{\mathrm{T}}$ can be extended to a continuous functional with domain

$$\mathcal{X}_{\mathrm{T}} = \overline{(\mathcal{S}_2, \|\mathrm{T}^{-1}\{\cdot\}\|_{\mathcal{X}})}, \tag{4.4}$$

which is identified as a predual of $\mathcal{X}'_{\mathrm{T}}$ since T^{-1} is continuous. Likewise the operator T^{-1} can be extended to a continuous and surjective operator $\overline{\mathrm{T}^{-1}}\colon \mathcal{X}_{\mathrm{T}} \to \mathcal{X}$. Now, it holds that $\langle f, \varphi\rangle = \langle g, \overline{\mathrm{T}^{-1}}\{\varphi\}\rangle$ for any $f = \mathrm{T}^{-*}\{g\}$ and $\varphi \in \mathcal{X}_{\mathrm{T}}$. Hence, the weak* convergence of a sequence $f_n = \mathrm{T}^{-*}\{g_n\}$ to $f = \mathrm{T}^{-*}\{g\}$ is equivalent to the weak* convergence of g_n to g.

Let us now discuss in more detail the case $\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$. To simplify the notation, we stick to $\mathcal{X} = C_0(\mathbb{R}^d)$, but the same argumentation applies to $L_p(\mathbb{R}^d)$. For the remainder of this section, we use the more specific notations

$$\mathcal{S}'_{\mathrm{Liz,T}}(\mathbb{R}^d) = \{\mathrm{T}^{-*}\{\mu\} \in \mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d) : \mu \in \mathcal{M}(\mathbb{R}^d)\} \tag{4.5}$$

and $\mathcal{S}_{\mathrm{Liz,T}}(\mathbb{R}^d)$ for the dual and for the predual, respectively. At first glance, the search for representatives for $\mathcal{S}'_{\mathrm{Liz,T}}(\mathbb{R}^d)$ is as difficult as before because we are still dealing with elements in $\mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$. To resolve this issue, let us assume that there exist continuous elements $\rho_{\mathrm{T},\mathbf{y}} = \mathrm{T}^{-*}\{\delta(\cdot - \mathbf{y})\} \in C(\mathbb{R}^d)$ in each equivalence class. Then, we conclude, for any $f = \mathrm{T}^{-*}\{\mu\} \in \mathcal{S}'_{\mathrm{Liz,T}}(\mathbb{R}^d)$ and $\varphi \in \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$, that

$$\begin{aligned}\langle \mathrm{T}^{-*}\{\mu\}, \varphi\rangle &= \langle \mu, \mathrm{T}^{-1}\{\varphi\}\rangle = \int_{\mathbb{R}^d} \mathrm{T}^{-1}\{\varphi\}(\mathbf{y})\mathrm{d}\mu(\mathbf{y}) = \int_{\mathbb{R}^d} \langle \mathrm{T}^{-*}\{\delta(\cdot - \mathbf{y})\}, \varphi\rangle \mathrm{d}\mu(\mathbf{y}) \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} (\rho_{\mathrm{T},\mathbf{y}}(\mathbf{x}) - p_{\mathbf{y}}(\mathbf{x}))\varphi(\mathbf{x})\mathrm{d}\mathbf{x}\mathrm{d}\mu(\mathbf{y}),\end{aligned} \tag{4.6}$$

where the polynomials $p_{\mathbf{y}}(\mathbf{x}) \in \mathcal{P}(\mathbb{R}^d)$ are added to ensure the following properties: First, the kernel $h(\mathbf{x},\mathbf{y}) = \rho_{\mathrm{T},\mathbf{y}}(\mathbf{x}) - p_{\mathbf{y}}(\mathbf{x})$ must be bi-continuous and bounded for some $g \in L_{\infty,\alpha}(\mathbb{R}^d)$, $\alpha \geq 0$, and every $\mathbf{y} \in \mathbb{R}^d$ by

$$|h(\mathbf{x},\mathbf{y})| \leq g(\|\mathbf{x}\|). \tag{4.7}$$

Second, we require that $h(\mathbf{x},\mathbf{y}) \to 0$ for a given $\mathbf{x} \in \mathbb{R}^d$ and $\|\mathbf{y}\| \to \infty$. Using Fubini's theorem and the growth control, we can then identify the distribution $\mathrm{T}^{-*}\{\mu\} \in \mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$ as the continuous function $f(\mathbf{x}) = \int_{\mathbb{R}^d} \rho_{\mathrm{T},\mathbf{y}}(\mathbf{x}) - p_{\mathbf{y}}(\mathbf{x})\mathrm{d}\mu(\mathbf{y})$. Due to the growth bound, this function corresponds to a unique distribution in $\mathcal{S}'(\mathbb{R}^d)$. We collect these observations together with a few properties in Theorem 4.1.

Theorem 4.1. *Assume that the Schwartz kernel $h(\mathbf{x},\mathbf{y}) = \rho_{\mathrm{T},\mathbf{y}}(\mathbf{x}) - p_{\mathbf{y}}(\mathbf{x})$ is bi-continuous and bounded for every $\mathbf{y} \in \mathbb{R}^d$ as*

$$|h(\mathbf{x},\mathbf{y})| \leq g(\|\mathbf{x}\|) \tag{4.8}$$

for some $g \in L_{\infty,\alpha}(\mathbb{R}^d)$, $\alpha \geq 0$. Then, any element $f = \mathrm{T}^{-*}\{\mu\} \in \mathcal{S}'_{\mathrm{Liz,T}}(\mathbb{R}^d)$ can be identified as the continuous function

$$f(\mathbf{x}) = \int_{\mathbb{R}^d} \rho_{\mathrm{T},\mathbf{y}}(\mathbf{x}) - p_{\mathbf{y}}(\mathbf{x})\mathrm{d}\mu(\mathbf{y}) \tag{4.9}$$

with bounded growth so that $|f(\mathbf{x})| \leq |\mu|(\mathbb{R}^d)g(\|\mathbf{x}\|)$. In particular, we get that the operator $\mathrm{P}_{\mathrm{Liz,T}} \colon \mathcal{S}'_{\mathrm{Liz,T}}(\mathbb{R}^d) \to L_{\infty,\alpha}(\mathbb{R}^d) \hookrightarrow \mathcal{S}'(\mathbb{R}^d)$ with $\mathrm{T}^{-*}\{\mu\} \mapsto f$ assigning the representatives is linear and continuous. Moreover, if $h(\mathbf{x},\cdot) \in C_0(\mathbb{R}^d)$ for every $\mathbf{x} \in \mathbb{R}^d$, then the point evaluations for the representatives are in the predual $\mathcal{S}_{\mathrm{Liz,T}}(\mathbb{R}^d)$.

Proof. We have already shown that f is indeed a representative. The growth bound and the continuity of $\mathrm{P}_{\mathrm{Liz,T}}$ follow immediately from the bound on $h(\mathbf{x},\mathbf{y})$.

Next, we show that point evaluations are weak*-continuous. Let $f_n = \mathrm{T}^{-*}\{\mu_n\}$ be a weak*-convergent sequence with limit $f = \mathrm{T}^{-*}\{\mu\}$, in the sense that μ_n converges weakly to μ as a measure. Due to the requirement that $e(\mathbf{x},\cdot) \in C_0(\mathbb{R}^d)$, we directly get that the evaluation functionals are weak*-continuous. To conclude the argument, we recall that the only weak*-continuous linear functionals on $\mathcal{S}'_{\mathrm{Liz},\mathrm{T}}(\mathbb{R}^d)$ are the elements of $\mathcal{S}_{\mathrm{Liz},\mathrm{T}}(\mathbb{R}^d)$, see [25, Theorem IV.20]. □

While this construction does not cover all Lizorkin distributions, we discuss two interesting examples, which will then be used in Sec. 5.2 to revisit representer theorems for certain problems.

Remark 4.1. The same argumentations and constructions can be applied if $\mathcal{S}_1$ consists of even or odd (hyper-spherical) Lizorkin functions. This setting is actually required for one of our examples.

4.1. *Example* 1: *Fractional Laplacians*

Here, we choose the spaces as $\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ and $\mathcal{X} = C_0(\mathbb{R}^d)$ with T being the fractional Laplacian described in Sec. 2, which is self-adjoint. Specifically, $(-\Delta)^\alpha \colon \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d) \to \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ for any $\alpha \in \mathbb{R}$ with $(-\Delta)^{-\alpha}(-\Delta)^\alpha = \mathrm{Id}$ on $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$. First, we note that the required density result was already established in Theorem 3.2. According to these choices, $\mathcal{X}'_{\mathrm{T}}$ is given by

$$\mathcal{M}^\alpha(\mathbb{R}^d) = \{(-\Delta)^{-\alpha/2}\mu \in \mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d) : \mu \in \mathcal{M}(\mathbb{R}^d)\}, \tag{4.10}$$

with predual space $C^\alpha(\mathbb{R}^d) = \overline{(\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d), \|(-\Delta)^{-\alpha/2}\cdot\|_{L_\infty})}$.

By Theorem 4.1, we can get a representation operator for $\alpha > d$ and $(\alpha - d) \notin \mathbb{N}$. Indeed, let $\rho_{\mathrm{Liz},\alpha} = (-\Delta)^{-\alpha/2}\{\delta\} \in \mathcal{M}^\alpha(\mathbb{R}^d)$ with a continuous representation given by (2.8). Using this representation, we obtain that $\rho_{\mathrm{Liz},\alpha}(\cdot - \mathbf{x}_k) = (-\Delta)^{-\alpha/2}\{\delta(\cdot - \mathbf{x}_k)\}$. Now, we have to show that there exist polynomials $p_{\mathbf{y}}(\mathbf{x}) \in \mathcal{P}(\mathbb{R}^d)$ such that the kernel $h(\mathbf{x},\mathbf{y}) = \rho_{\mathrm{Liz},\alpha}(\mathbf{x}-\mathbf{y}) - p_{\mathbf{y}}(\mathbf{x})$ fulfills the requirements. Based on (2.8), we construct for $\mathbf{y} \neq \mathbf{0}$ the polynomial $\tilde{p}_{\mathbf{y}}(\mathbf{x}) = T_{\lceil\alpha-d-1\rceil}\{\rho_{\mathrm{Liz},\alpha}(\cdot - \mathbf{y})\}(\mathbf{x})$ with $T_{\lceil\alpha-d-1\rceil}$ the Maclaurin expansion of order $\lceil\alpha - d - 1\rceil$ around $\mathbf{0}$. For this function, we can bound the kernel $\tilde{h}(\mathbf{x},\mathbf{y}) = \rho_{\mathrm{Liz},\alpha}(\mathbf{x}-\mathbf{y}) - \tilde{p}_{\mathbf{y}}(\mathbf{x})$ by

$$|\tilde{h}(\mathbf{x},\mathbf{y})| \le C\|\mathbf{x}\|^{\lceil\alpha-d\rceil} \sup_{t\in[0,1],|\mathbf{k}|=\lceil\alpha-d\rceil} \|\partial^{\mathbf{k}}\rho_{\mathrm{Liz},\alpha}(t\mathbf{x}-\mathbf{y})\|_2. \tag{4.11}$$

For any fixed $\mathbf{x} \in \mathbb{R}^d$, we then use our estimates from Proposition A.1 in the appendix to conclude that $|\tilde{h}(\mathbf{x},\mathbf{y})| \le C\|\mathbf{x}\|^{\alpha-d}$ if $\|\mathbf{y}\| \ge \|\mathbf{x}\| + 1$ and $\tilde{e}(\mathbf{x},\cdot) \in C_0(\mathbb{R}^d)$ for every $\mathbf{x} \in \mathbb{R}^d$. Next, apply a smooth function $\chi\colon \mathbb{R} \to \mathbb{R}$ with $\chi(t) = 0$ if $|t| \le 1$ and $\chi(t) = 1$ if $|t| \ge 2$ to define a bi-continuous (both in $\mathbf{x}$ and $\mathbf{y}$) function $x \mapsto p_{\mathbf{y}}(\mathbf{x}) = \chi(\|\mathbf{y}\|)\tilde{p}_{\mathbf{y}}(\mathbf{x}) \in \mathcal{P}(\mathbb{R}^d)$. Now, we can bound the kernel $h(\mathbf{x},\mathbf{y})$ using the bound for $\tilde{h}(\mathbf{x},\mathbf{y})$ by

$$|h(\mathbf{x},\mathbf{y})| \leq \max\left\{\max_{\|\mathbf{y}\|\leq\|\mathbf{x}\|+2}\{\rho_{\mathrm{Liz},\alpha}(\mathbf{x}-\mathbf{y})+|p_{\mathbf{y}}(\mathbf{x})|\}, C\|\mathbf{x}\|^{\alpha-d}\right\}$$
$$\leq C(\|\mathbf{x}\|+2)^{\alpha-d}, \tag{4.12}$$

where we used Proposition A.1 to produce the estimate

$$\max_{\|\mathbf{y}\|\leq\|\mathbf{x}\|+2}|p_{\mathbf{y}}(\mathbf{x})| \leq \max_{\|\mathbf{y}\|\leq\|\mathbf{x}\|+2}\sum_{|\mathbf{k}|\leq\lceil\alpha-d-1\rceil}\frac{1}{\mathbf{k}!}|\partial^{\mathbf{k}}\rho_{\mathrm{Liz},\alpha}(-\mathbf{y})|\|\mathbf{x}\|^{|\mathbf{k}|}$$
$$\leq C(\|\mathbf{x}\|+2)^{\alpha-d}. \tag{4.13}$$

Finally, we apply Theorem 4.1 to obtain the desired representations. The results of this section are summarized in the following corollary.

Corollary 4.1. *Let $\alpha > d$ with $(\alpha - d) \notin \mathbb{N}$. Then, there exists a continuous representation operator $\mathrm{P}_{\mathrm{Liz},\alpha}\colon \mathcal{M}^{\alpha}(\mathbb{R}^d) \to L_{\infty,\alpha-d}(\mathbb{R}^d) \subset \mathcal{S}'(\mathbb{R}^d)$ with $(-\Delta)^{-\alpha/2}\{\mu\} \mapsto f$. Further, the point evaluations for these representatives are in the predual $C^{\alpha}(\mathbb{R}^d)$.*

4.2. *Example 2: Radon domain splines*

In this example, we work with certain hyper-spherical counterparts of $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ as described in Sec. 2. More specifically, the Euclidean indexing with $\mathbf{x} \in \mathbb{R}^d$ is replaced by $(t,\boldsymbol{\xi}) \in \mathbb{R} \times \mathbb{S}^{d-1}$ and we distinguish between even and odd functions. We express this distinction with an index $m \in \mathbb{N}$, which simplifies the notation. We define

$$\mathcal{S}_{\mathrm{Liz},m}(\mathbb{R}\times\mathbb{S}^{d-1})$$
$$= \left\{\varphi \in \mathcal{S}_m(\mathbb{R}\times\mathbb{S}^{d-1}) : \int_{\mathbb{R}\times\mathbb{S}^{d-1}} \varphi(t,\boldsymbol{\xi})p(t)\mathrm{d}t\mathrm{d}\boldsymbol{\xi} = 0 \ \forall\, p \in \mathcal{P}(\mathbb{R})\right\}, \tag{4.14}$$

where $\mathrm{d}\boldsymbol{\xi}$ stands for the surface element on the unit sphere $\mathbb{S}^{d-1}$. Here, the space $\mathcal{S}_m(\mathbb{R}\times\mathbb{S}^{d-1})$ is defined as the even functions in $\mathcal{S}(\mathbb{R}\times\mathbb{S}^{d-1})$ if m is even and the odd ones otherwise. Correspondingly, an element $g \in \mathcal{S}'_{\mathrm{Liz},m}(\mathbb{R}\times\mathbb{S}^{d-1})$ is a continuous linear functional on $\mathcal{S}_{\mathrm{Liz},m}(\mathbb{R}\times\mathbb{S}^{d-1})$ whose action on the test function ϕ is represented by the duality product $\langle g,\phi\rangle_{\mathrm{Rad}}$. If g can be identified with a function $g\colon \mathbb{R}\times\mathbb{S}^{d-1} \to \mathbb{R}$, then

$$\langle g,\phi\rangle_{\mathrm{Rad}} = \int_{\mathbb{S}^{d-1}}\int_{\mathbb{R}} g(t,\boldsymbol{\xi})\phi(t,\boldsymbol{\xi})\mathrm{d}t\mathrm{d}\boldsymbol{\xi}. \tag{4.15}$$

The evaluation functional on $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}\times\mathbb{S}^{d-1})$ is $\delta_{\mathbf{z}_0} = \delta(\cdot - t_0)\delta(\cdot - \boldsymbol{\xi}_0)$ with $\mathbf{z}_0 = (t_0,\boldsymbol{\xi}_0) \in \mathbb{R}\times\mathbb{S}^{d-1}$. A brief overview for properties of the Radon transform R and its filtered version $\mathrm{K}_{\mathrm{rad}}\mathrm{R}$ related to these spaces is given in Appendix B. In particular, it holds that both R and $\mathrm{K}_{\mathrm{rad}}\mathrm{R}$ are homeomorphisms. Next, we briefly review Lizorkin ridges, which play a key role for the construction of representatives.

Lizorkin Ridges The 1D profile (or ridge) along the direction $\boldsymbol{\xi}_0 \in \mathbb{R}^d$ associated to $r \in \mathcal{S}'_{\mathrm{Liz}}(\mathbb{R})$ is the distribution $r_{\boldsymbol{\xi}_0} \in \mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$ that satisfies

$$\forall\, \varphi \in \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d) : \langle r_{\boldsymbol{\xi}_0}, \varphi\rangle = \langle r, \mathrm{R}\{\varphi\}(\cdot, \boldsymbol{\xi}_0)\rangle. \tag{4.16}$$

The most basic ridge is $\delta(\boldsymbol{\xi}_0^\mathsf{T} \cdot - t_0) \coloneqq r_{\boldsymbol{\xi}_0}$ with $r = \delta(\cdot - t_0)$. It is a Dirac ridge along $\boldsymbol{\xi}_0$ with offset t_0. Since the Fourier transform of such ridges is localized along the ray $\{\boldsymbol{\omega} = \omega\boldsymbol{\xi}_0 : \omega \in \mathbb{R}\}$, the Radon transform of a ridge must vanish away from $\pm\boldsymbol{\xi}_0$. This is generalized and formalized as follows.

Proposition 4.1 (Radon Transform of Lizorkin Ridges). *Let* $(t_0, \boldsymbol{\xi}_0) = \mathbf{z}_0 \in \mathbb{R} \times \mathbb{S}^{d-1}$ *and* $r \in \mathcal{S}'_{\mathrm{Liz}}(\mathbb{R})$*. Then,*

$$\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{\delta(\boldsymbol{\xi}_0^\mathsf{T} \cdot - t_0)\} = \mathrm{P}_{\mathrm{even}}\{\delta_{\mathbf{z}_0}\} \in \mathcal{S}'_{\mathrm{Liz}}(\mathbb{R} \times \mathbb{S}^{d-1}), \tag{4.17}$$

$$\mathrm{R}\{\delta(\boldsymbol{\xi}_0^\mathsf{T} \cdot)\} = \mathrm{P}_{\mathrm{even}}\{q_d\delta(\cdot - \boldsymbol{\xi}_0)\} \in \mathcal{S}'_{\mathrm{Liz}}(\mathbb{R} \times \mathbb{S}^{d-1}), \tag{4.18}$$

$$\mathrm{K}_{\mathrm{rad}}\mathrm{R}\{r(\boldsymbol{\xi}_0^\mathsf{T} \cdot)\} = \mathrm{P}_{\mathrm{even}}\{r\delta(\cdot - \boldsymbol{\xi}_0)\} \in \mathcal{S}'_{\mathrm{Liz}}(\mathbb{R} \times \mathbb{S}^{d-1}), \tag{4.19}$$

$$\mathrm{R}\{r(\boldsymbol{\xi}_0^\mathsf{T} \cdot)\} = \mathrm{P}_{\mathrm{even}}\{(q_d * r)\delta(\cdot - \boldsymbol{\xi}_0)\} \in \mathcal{S}'_{\mathrm{Liz}}(\mathbb{R} \times \mathbb{S}^{d-1}), \tag{4.20}$$

where $q_d(t) = 2(2\pi)^{d-1}\mathcal{F}^{-1}\{1/|\cdot|^{d-1}\}(t)$ *is the 1D impulse response of the Radon-domain inverse filtering operator* $\mathrm{K}_{\mathrm{rad}}^{-1}$*. Here,* (4.17) *can be identified as an even measure.*

Proof. For any $\varphi \in \mathcal{S}_{\mathrm{Liz},0}(\mathbb{R} \times \mathbb{S}^{d-1})$, it holds that $\mathrm{R}\mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\{\varphi\} = \varphi$ and therefore, also that

$$\langle \mathrm{K}_{\mathrm{rad}}\mathrm{R}\{r_{\boldsymbol{\xi}_0}\}, \varphi\rangle = \langle r_{\boldsymbol{\xi}_0}, \mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\{\varphi\}\rangle = \langle r, \mathrm{R}\mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\{\varphi\}(\cdot, \boldsymbol{\xi}_0)\rangle = \langle r, \varphi(\cdot, \boldsymbol{\xi}_0)\rangle, \tag{4.21}$$

from which (4.17) and (4.19) do follow. In a similar way, we obtain, for any $\varphi \in \mathcal{S}_{\mathrm{Liz},0}(\mathbb{R} \times \mathbb{S}^{d-1})$, that

$$\langle \mathrm{R}\{r_{\boldsymbol{\xi}_0}\}, \varphi\rangle = \langle r, \mathrm{R}\mathrm{R}^*\{\varphi\}(\cdot, \boldsymbol{\xi}_0)\rangle = \langle r, \mathrm{K}_{\mathrm{rad}}^{-1}\{\varphi\}(\cdot, \boldsymbol{\xi}_0)\rangle, \tag{4.22}$$

from which (4.18) and (4.20) do follow as $\mathrm{K}_{\mathrm{rad}}^{-1}\{\varphi\}(t, \boldsymbol{\xi}_0) = (q_d * \varphi(\cdot, \boldsymbol{\xi}_0))(t)$. □

An equivalent form of (4.17) in Proposition 4.1 is

$$\delta(\boldsymbol{\xi}_0^\mathsf{T} \cdot - t_0) = \mathrm{R}^*\mathrm{P}_{\mathrm{even}}\{\delta_{\mathbf{z}_0}\}(\mathbf{x}), \tag{4.23}$$

which results from $\mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\mathrm{R} = \mathrm{Id}$ on $\mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$. Note that the other identities can be rewritten in a similar form, too.

Construction of Radon Splines In this example, we choose the spaces for constructing the Banach subspaces as $\mathcal{S}_1 = \mathcal{S}_{\mathrm{Liz},m}(\mathbb{R} \times \mathbb{S}^{d-1})$, $\mathcal{S}_2 = \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ and $\mathcal{X} = C_{0,m}(\mathbb{R}^d)$, where $C_{0,m}(\mathbb{R}^d)$ consists of even or odd continuous functions, respectively, that vanish at infinity. Next, recall that the derivative $\partial_t^m : \mathcal{S}_{\mathrm{Liz},m}(\mathbb{R} \times \mathbb{S}^{d-1}) \to \mathcal{S}_{\mathrm{Liz},0}(\mathbb{R} \times \mathbb{S}^{d-1})$ is self-adjoint and a homeomorphism. Its inverse can be constructed by iterating $\partial_t^{-1}\{\varphi\}(t, \boldsymbol{\xi}) = \int_t^\infty \varphi(r, \boldsymbol{\xi})\mathrm{d}r$. Then, we choose

$\mathrm{T} = \mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\partial_t^m$ such that the dual $\mathrm{T}^* = \partial_t^m \mathrm{K}_{\mathrm{rad}}\mathrm{R}$ is the concatenation of the filtered projection $\mathrm{K}_{\mathrm{rad}}\mathrm{R}\colon \mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d) \to \mathcal{S}'_{\mathrm{Liz},0}(\mathbb{R} \times \mathbb{S}^{d-1})$ with the partial derivative ∂_t^m. To begin, we show the required density result for the construction of the spaces related to Theorem 4.1, which also applies in this hyper-spherical setting, as pointed out in Remark 4.1.

Lemma 4.1. *It holds that*

$$\overline{(\mathcal{S}_{\mathrm{Liz},m}(\mathbb{R}^d), \|\cdot\|_\infty)} = C_{0,m}(\mathbb{R}^d) = \begin{cases} C_{0,\mathrm{even}}(\mathbb{R}^d) & \text{if } m \text{ is even,} \\ C_{0,\mathrm{odd}}(\mathbb{R}^d) & \text{if } m \text{ is odd.} \end{cases} \tag{4.24}$$

Proof. By the Stone–Weierstrass theorem and the continuity of the projection onto even or odd functions, respectively, we first get that $\overline{C_{0,m}(\mathbb{R}) \times C(\mathbb{S}^{d-1})} = C_{0,m}(\mathbb{R}^d)$. Then, we conclude from Theorem 3.2 that $\mathcal{S}_{\mathrm{Liz,m}}(\mathbb{R}) \times C^\infty(\mathbb{S}^{d-1}) \subset \mathcal{S}_{\mathrm{Liz,m}}(\mathbb{R} \times \mathbb{S}^{d-1})$ is dense in $C_{0,m}(\mathbb{R}^d)$. □

According to these choices, our Banach space $\mathcal{X}'_{\mathrm{T}}$ with smoothness exponent m is given by

$$\mathcal{M}_{\mathrm{Rad},m}(\mathbb{R}^d) = \{\mathrm{R}^*\partial_t^{-m}\{\mu\} \in \mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d) : \mu \in \mathcal{M}_m(\mathbb{R}^d)\} \tag{4.25}$$

with predual $C_{\mathrm{Rad},m}(\mathbb{R}^d) = \overline{(\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d), \|\partial_t^{-m}\mathrm{R}\{\cdot\}\|_{L_\infty})}$. Now, we show that Theorem 4.1 can be applied for $m \geq 2$ to get continuous representations of elements in $\mathcal{M}_{\mathrm{Rad},m}(\mathbb{R}^d)$. Define

$$\rho_{\mathrm{Rad},m}(x) = \max(0,x)^{m-1}/(m-1)!. \tag{4.26}$$

Then, as shown in Proposition 4.1, $\rho_{\mathrm{Rad},m}(\langle\cdot,\boldsymbol{\xi}_0\rangle - t_0)$ with $\mathbf{z}_0 = (t_0, \boldsymbol{\xi}_0) \in \mathbb{R} \times \mathbb{S}^{d-1}$ is an element of the equivalence class

$$\frac{1}{2}\mathrm{R}^*\mathrm{P}_{\mathrm{even}}\{\rho_{\mathrm{Rad},m}(\cdot - t_0)\delta(\cdot - \boldsymbol{\xi}_0)\} = \frac{1}{2}\mathrm{R}^*\partial_t^{-m}\{\delta_{\mathbf{z}_0} \pm \delta_{-\mathbf{z}_0}\} \in \mathcal{M}_{\mathrm{Rad},m}(\mathbb{R}^d), \tag{4.27}$$

where the sign depends on m. Now, we have to show that there are polynomials $p_{t,\boldsymbol{\xi}} \in \mathcal{P}(\mathbb{R}^d)$ such that the kernel $h(\mathbf{x},\mathbf{z}) = \rho_{\mathrm{Rad},m}(\langle\mathbf{x},\boldsymbol{\xi}\rangle - t) - p_{t,\boldsymbol{\xi}}(\mathbf{x})$ with $z = (t,\boldsymbol{\xi})$ fulfills the requirements. As m is a natural number, no Taylor expansion is necessary and we can provide the correcting family of polynomials directly. More precisely, we set

$$p_{t,\boldsymbol{\xi}} = \max\{0, \min\{-t, 1\}\}(\langle\cdot,\boldsymbol{\xi}\rangle - t)^{m-1}/(m-1)! \in \mathcal{P}(\mathbb{R}^d), \tag{4.28}$$

which ensures that $(\rho_{\mathrm{Rad},m}(\langle\mathbf{x},\cdot\rangle - \cdot) - p_{\{\cdot\}}(\mathbf{x})) \in C_0(\mathbb{R} \times \mathbb{S}^{d-1})$ together with

$$\|\rho_{\mathrm{Rad},m}(\langle\mathbf{x},\cdot\rangle - \cdot) - p_{\{\cdot\}}(\mathbf{x})\|_\infty \leq C\|\mathbf{x}\|^{m-1}. \tag{4.29}$$

Hence, we can apply Theorem 4.1 to obtain explicit representations.

Corollary 4.2. *Let $m \geq 2$. Then, there exists a continuous representation operator $\mathrm{P}_{\mathrm{Rad},m}\colon \mathcal{M}_{\mathrm{Rad},m}(\mathbb{R}^d) \to L_{\infty,m-1}(\mathbb{R}^d) \subset \mathcal{S}'(\mathbb{R}^d)$ with $\mathrm{R}^*\partial_t^{-m}\{\mu\} \mapsto f$. Further, the point evaluations for these representatives are in the predual $C_{\mathrm{Rad},m}(\mathbb{R}^d)$.*

5. Variational Problems that Involve Lizorkin Spaces

As warm-up, we first revisit periodic (fractional) splines [9]. For variational problems in which the regularization favors such functions, we can use the projection (3.7) to get representations. Note that our approach is applicable to a very broad class of problems; namely, whenever a continuous projection and a suitable extension are available. By contrast, after this warm-up example, we study variational problems where no projection onto the involved spaces is available. There, we focus on problems that involve our previously constructed Banach subspaces, which usually only consist of equivalence classes. This makes the situation much more delicate than before and the use of a representation operator is necessary. Based on this operator, we are able to obtain similar results as before.

5.1. *Periodic fractional splines*

We use our tools to derive a representer theorem that is an alternative to the one in [9]. To this end, we need the space $C(\mathbb{T}) = \overline{(\mathcal{S}(\mathbb{T}), \|\cdot\|_{L_\infty})}$ of continuous, T-periodic functions. Its topological dual $\mathcal{M}(\mathbb{T})$ (namely, the space of T-periodic Radon measures) can be specified as

$$\mathcal{M}(\mathbb{T}) = \{f \in \mathcal{S}'(\mathbb{T}) : \|f\|_{\mathcal{M}} < \infty\} \quad \text{with } \|f\|_{\mathcal{M}} \coloneqq \sup_{\phi \in \mathcal{S}(\mathbb{T}): \|\phi\|_{L_\infty} \leq 1} \langle f, \phi \rangle. \tag{5.1}$$

Since the projection (3.7) continuously extends to these spaces, we have the decomposition $C(\mathbb{T}) = C_0(\mathbb{T}) \oplus \mathcal{P}_0$ with $C_0(\mathbb{T}) = \mathrm{P}_0(C(\mathbb{T}))$. The final ingredient are the sampling functionals in $\mathcal{M}_0(\mathbb{T}) = C_0(\mathbb{T})' \simeq \mathcal{M}(\mathbb{T})/\mathcal{P}_0$, where we use P_0^* to identify representations.

Theorem 5.1 (Periodic Lizorkin Sampling Functionals). *The Lizorkin sampling functionals* $\delta_0(\cdot - t_0) = \mathrm{P}_0^*\{\delta_{\mathrm{perio}}(\cdot - t_0)\} \in \mathcal{M}_0(\mathbb{T})$ *with* $t_0 \in \mathbb{T}$ *have the following properties*:

(1) *Explicit representation*: $\delta_0(\cdot - t_0) = \delta_{\mathrm{perio}}(\cdot - t_0) - 1$.
(2) *Sampling at* $t_0 : \langle \delta_0(\cdot - t_0), \phi \rangle = \phi(t_0)$ *for all* $\phi \in C_0(\mathbb{T})$.
(3) *Zero mean*: $\langle \delta_0(\cdot - t_0), 1 \rangle = 0$ *for all* $t_0 \in \mathbb{R}$.
(4) *It holds that* $\|\delta_0(\cdot - t_0)\|_{\mathcal{M}_0} = 1$ *for any* $t_0 \in \mathbb{T}$.
(5) *For finite sets* $\{t_k\} \subset \mathbb{T}$ *of distinct points, it holds* $\|\sum_k a_k \delta_0(\cdot - t_k)\|_{\mathcal{M}_0} = \sum_k |a_k|$.
(6) *If* $e_k \in \mathrm{Ext}B(\mathcal{M}_0(\mathbb{T}))$, *then* $e_k = \pm\delta_0(\cdot - t_k)$ *for some* $t_k \in \mathbb{T}$.

Proof. The first 3 items follow directly by construction. Now, we prove item 4. For any $(f, \phi) \in \mathcal{M}(\mathbb{T}) \times C_0(\mathbb{T})$, it holds that

$$\langle \mathrm{P}_0^*\{f\}, \phi \rangle = \langle f, \mathrm{P}_0\{\phi\} \rangle = \langle f, \phi \rangle. \tag{5.2}$$

In particular, $\langle \delta_0(\cdot - t_0), \phi\rangle = \langle \delta_{\text{perio}}(\cdot - t_0), \phi\rangle$. By definition of the dual norm, we have that

$$\begin{aligned}\|\delta_0(\cdot - t_0)\|_{\mathcal{M}_0} &= \sup_{\phi \in C_0(\mathbb{T}): \|\phi\|_{L_\infty} \leq 1} \langle \delta_0(\cdot - t_0), \phi\rangle = \sup_{\phi \in C_0(\mathbb{T}): \|\phi\|_{L_\infty} \leq 1} \langle \delta_{\text{perio}}(\cdot - t_0), \phi\rangle \\ &\leq \sup_{\phi \in C(\mathbb{T}): \|\phi\|_{L_\infty} \leq 1} \langle \delta_{\text{perio}}(\cdot - t_0), \phi\rangle = \|\delta_{\text{perio}}(\cdot - t_0)\|_{\mathcal{M}} = 1. \qquad (5.3)\end{aligned}$$

Next, we show that this bound is sharp by fixing $0 < \epsilon < T/2$ and choosing the test function

$$\phi_{\epsilon,\text{perio}}(\cdot - t_0) = \sum_{n \in \mathbb{Z}} \varphi_0\left(\frac{\cdot + nT - t_0}{\epsilon}\right), \qquad (5.4)$$

where $\varphi_0 : \mathbb{R} \to [-1, 1]$ is continuous with $\varphi_0(0) = 1$, $\int_{\mathbb{R}} \varphi_0(t)\mathrm{d}t = 0$ and $\text{supp}(\varphi_0) \subset [-1, 1]$. Then, the statement follows from $\phi_{\epsilon,\text{perio}}(t_0) = 1$ and $\|\phi_{\epsilon,\text{perio}}\|_{L_\infty} \leq 1$. Similarly, for item 5, we first observe that the triangle inequality leads to

$$\sup_{\phi \in C_0(\mathbb{T}): \|\phi\|_{L_\infty} \leq 1} \left\langle \sum_k a_k \delta_0(\cdot - t_k), \phi \right\rangle = \left\| \sum_k a_k \delta_0(\cdot - t_k) \right\|_{\mathcal{M}_0} \leq \sum_k |a_k|. \qquad (5.5)$$

Since the t_k are distinct, there exists $\epsilon > 0$ with $|t_k - t_{k'}| > 2\epsilon$ for all $k' \neq k$. Then, we take the critical function $\phi_{\text{crit}}(t) = \sum_k \text{sgn}(a_k)\phi_{\epsilon,\text{perio}}(t - t_k)$ satisfying $\|\phi_{\text{crit}}\|_{\mathrm{L}_\infty} = 1$, which saturates the bound.

Due to $M_0(\mathbb{T}) \simeq \mathcal{M}(\mathbb{T})/\mathcal{P}_0$, it holds that $\mathrm{P}_0^* B(\mathcal{M}(\mathbb{T})) = B(\mathcal{M}_0(\mathbb{T}))$. By [4, Lemma 3.2], we then get that $\text{Ext}B(\mathcal{M}_0(\mathbb{T})) \subset \mathrm{P}_0^*\text{Ext}B(\mathcal{M}(\mathbb{T}))$. Since the extreme points of $B(\mathcal{M}(\mathbb{T}))$ are $\{\pm\delta(\cdot - t)\}_{t \in \mathbb{T}}$, the last claim readily follows. □

Now, we are able to formulate the approximation problem. Given a series of (possibly noisy) data points $(y_m, t_m) \in \mathbb{R} \times \mathbb{T}$, $m = 1, \dots, M$, we consider the task of reconstructing a periodic function $f : \mathbb{T} \to \mathbb{R}$ such that $f(t_1) \approx y_1, \dots, f(t_M) \approx y_M$ without overfitting. Since this problem is inherently ill-posed, we put a penalty on $\|\mathrm{D}^\alpha\{f\}\|_{\mathcal{M}_0}$ in order to favor solutions with “sparse” αth derivatives. The corresponding native space is

$$\begin{aligned}\mathcal{M}^\alpha(\mathbb{T}) &= \{f \in \mathcal{S}'(\mathbb{T}) : \|\mathrm{D}^\alpha\{f\}\|_{\mathcal{M}_0} < \infty\} \\ &= \{\mathrm{D}^{-\alpha}\{w\} + p_0 : (w, p_0) \in \mathcal{M}_0(\mathbb{T}) \times \mathcal{P}_0\}. \qquad (5.6)\end{aligned}$$

In particular, this means that $\mathcal{M}^\alpha(\mathbb{T}) = \mathcal{U}' \oplus \mathcal{P}_0$ with $\mathcal{U}' = \mathrm{D}^{-\alpha}(\mathcal{M}_0(\mathbb{T}))$, which is isomorphic to $\mathcal{M}_0(\mathbb{T}) \times \mathcal{P}_0$. The basic atoms for the representation of minimum-norm interpolators in $\mathcal{U}'$ are the extreme points e_k of the unit ball $B_{\mathcal{U}'}(1)$. Due to the isometry between $\mathcal{U}'$ and $\mathcal{M}_0(\mathbb{T})$, we have that $\text{Ext}B_{\mathcal{U}'}(1) = \mathrm{D}^{-\alpha}(\text{Ext}B_{\mathcal{M}_0}(1))$, which in light of items 1 and 6 in Theorem 5.1 yields that

$$e_k = \mathrm{D}^{-\alpha}\{\delta_0(\cdot - t_k)\} = \rho_{\mathrm{perio},\alpha}(\cdot - t_k), \tag{5.7}$$

where

$$\rho_{\mathrm{perio},\alpha}(t) = \mathrm{D}^{-\alpha}\{\delta_0\}(t) = \sum_{n\in\mathbb{Z}\setminus\{0\}} \frac{1}{(\mathrm{j}n\omega_0)^\alpha}\mathrm{e}^{\mathrm{j}n\omega_0 t}. \tag{5.8}$$

The latter formula is obtained from (3.8) by using that $\widehat{\delta}_0[n] = \widehat{\delta}[n]$ for $n \neq 0$. The resulting Fourier series (5.8) converges to a continuous function for $\alpha > 1$. The functions $\rho_{\mathrm{perio},\alpha}$ are the building blocks of the (non-periodic) fractional splines of degree $\alpha - 1$. Now, the direct application of the third case of [38, Theorem 3] yields the following.

Theorem 5.2 (Minimum-Energy Periodic Spline Reconstruction). *Let $E\colon \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ be a strictly convex loss function and $\lambda > 0$ some regularization parameter. Then, for any given data points $(y_m, t_m) \in \mathbb{R}\times\mathbb{T}, m = 1,\dots,M$, the solution set of the functional-approximation problem with $\alpha > 1$,*

$$S = \operatorname*{arg\,min}_{f\in\mathcal{M}^\alpha(\mathbb{T})} \sum_{m=1}^{M} E(y_m, f(t_m)) + \lambda\|\mathrm{D}^\alpha f\|_{\mathcal{M}_0} \tag{5.9}$$

is nonempty and weak-compact. It is the weak* closure of the convex hull of its extreme points, which are all of the form*

$$f_{\mathrm{Ext}}(t) = b_0 + \sum_{k=1}^{K_0} a_k \rho_{\mathrm{perio},\alpha}(t - \tau_k) \tag{5.10}$$

for some $K_0 \leq M - 1$, weights and knots $(a_k, \tau_k) \in \mathbb{R}\times\mathbb{R}$, $k = 1,\dots,K_0$, and the periodic basis function $\rho_{\mathrm{perio},\alpha}\colon \mathbb{R}\to\mathbb{R}$ specified by (5.8).

Proof. First, we identify the (unique) predual space $C_0^\alpha(\mathbb{T}) = \mathcal{U} \oplus \mathcal{P}_0$ such that $\mathcal{M}^\alpha(\mathbb{T}) = \mathcal{U}' \oplus \mathcal{P}_0'$. By the injectivity of $\mathrm{D}^{\alpha*}$ on $C_0(\mathbb{T}) = \overline{(\mathcal{S}_0(\mathbb{R}), \|\cdot\|_{L_\infty})}$ and by setting $\mathcal{U} = \mathrm{D}^{\alpha*}(C_0(\mathbb{T}))$, we readily verify that $\mathcal{U}' = \mathrm{D}^{-\alpha}(\mathcal{M}_0(\mathbb{T}))$. This allows us to identify the predual space as

$$C_0^\alpha(\mathbb{T}) = \mathcal{U} \oplus \mathcal{P}_0 = \{\mathrm{D}^{\alpha*}\{v\} + p_0 : (v, p_0) \in C_0(\mathbb{T}) \times \mathcal{P}_0\}, \tag{5.11}$$

which is a Banach space isomorphic to $C_0(\mathbb{T}) \times \mathcal{P}_0$ as expected. The technical prerequisite for applying [38, Theorem 3] is the weak*-continuity of the sampling functionals $\delta(\cdot - t_m)$, which is equivalent to $\delta(\cdot - t_m) \in \mathcal{C}_0^\alpha(\mathbb{T})$. To this end, we have that

$$\mathrm{D}^{-\alpha*}\{\delta(\cdot - t_m) - 1\} = \rho_{\mathrm{perio},\alpha}(t_m - \cdot) \tag{5.12}$$

with the latter function being included in $C_0(\mathbb{T})$ if and only if $\alpha > 1$ or, equivalently, when the Fourier coefficients in (5.8) are in $\ell_1(\mathbb{Z})$. □

Remark 5.1. Functions of the form (5.10) are fractional splines if and only if $\sum_{k=1}^{K_0} a_k = 0$, see [9, Proposition 3]. To ensure this, we can add the constraint $\langle f, e^{\mathrm{j}\omega_0 \cdot}\rangle_{\mathbb{T}} = 0$ in Theorem 5.2, which again leads to extreme points of the form (5.10) with $K_0 \leq M$. Since $\|\mathrm{D}^\alpha f\|_{\mathcal{M}} \geq \|\mathrm{D}^\alpha f\|_{\mathcal{M}_0}$ with equality holding for extreme points, this modified version of Theorem 5.2 remains true if we replace $\|\mathrm{D}^\alpha f\|_{\mathcal{M}_0}$ in (5.9) by $\|\mathrm{D}^\alpha f\|_{\mathcal{M}}$. Plots of the fractional splines $\rho_{\mathrm{perio},\alpha} - \rho_{\mathrm{perio},\alpha}(\cdot - \frac{T}{2})$ for different α are given in [9, Fig. 1].

Remark 5.2 (Numerical Approach). To find f, we can overparameterize it with knots τ_k chosen over a fine uniform grid. The respective weights are then recovered by solving a discrete penalized basis pursuit problem using state-of-the-art proximal algorithms [7, 15] or Bregman methods [5]. While conceptually simple, this is computationally expensive since the underlying grid needs to have much more knots than $M - 1$. More advanced meshfree approaches for directly recovering the positions τ_k can be developed using, for example, the Franck–Wolfe algorithm [8, 12].

5.2. *A general variational problem framework*

In this section, we first state a general variational problem framework that involves the constructed Banach spaces and shares some similarities with the approach presented in Sec. 5.1, but for which no projector is available. Here, the derived representation operator from Theorem 4.1 makes the framework explicit, again with the advantage that we can rely on the general abstract machinery for the derivation of theoretical results. We then treat several useful special cases related with the Banach subspaces of $\mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$ introduced as examples in Sec. 4.

By construction, we immediately deduce that the extreme points of the unit ball in $\mathcal{X}'_{\mathrm{T}}$ are given by $\tilde{e}_k = \mathrm{T}^{-*}\{e_k\} \in \mathcal{X}'_{\mathrm{T}}$, where e_k are the extreme points of the unit ball in $\mathcal{X}'$. Now, we are able to formulate a variational problem that involves our constructed Banach spaces and provide a representer theorem for the structure of the solutions.

Theorem 5.3 (Representer Theorem [38]). *Let the linear operator* $\nu\colon \mathcal{X}'_{\mathrm{T}} \to \mathbb{R}^M$ *be given by* $f \mapsto (\langle \nu_1, f\rangle, \dots, \langle \nu_M, f\rangle)$ *with* $\nu_i \in \mathcal{X}_{\mathrm{T}}$ *being linearly independent. Further, let* $E\colon \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}_+ \cup \{+\infty\}$ *be proper, lower-semicontinuous and convex and let* $\psi\colon \mathbb{R}_+ \to \mathbb{R}_+$ *be strictly increasing and convex. Then, for any fixed* $y \in \mathbb{R}^M$, *the solution set* S *of the generic optimization problem*

$$\operatorname*{arg\,min}_{f \in \mathcal{X}'_{\mathrm{T}}} E(y, \nu\{f\}) + \psi(\|f\|_{\mathcal{X}'_{\mathrm{T}}}) \tag{5.13}$$

is nonempty, convex and weak-compact. If, additionally,* E *is strictly convex or if it imposes the equality constraint* $y = \nu\{f\}$, *then* S *is the weak* closure of the* *convex hull of its extreme points, which can all be expressed as*

$$f_0 = \sum_{k=1}^{K_0} c_k \mathrm{T}^{-*}\{e_k\} \tag{5.14}$$

with $K_0 \leq M$ *and* $c_k \in \mathbb{R}$.

Remark 5.3. The result can be slightly strengthened if $\mathcal{X}'$ is strictly convex, see [38] for details.

As illustration, we briefly derive two corollaries from Theorem 5.3. They are based on the two Banach subspaces of $\mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$ introduced in Sec. 4.

5.3. *Fractional splines*

Here, we extend our investigations in Sec. 5.1 to non-periodic splines using Theorem 5.3 and the discussion from Sec. 4.1, which is summarized in Corollary 4.1. Since the point evaluations are in the predual, the application of Theorem 5.3, together with the explicit representation of elements in $\mathcal{M}^\alpha(\mathbb{R}^d)$, yields the following.

Corollary 5.1 (Minimum-Energy Lizorkin Splines). *Let* $E\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ *be a strictly convex loss function, let* $(\mathbf{x}_m, y_m) \in \mathbb{R}^d \times \mathbb{R}, m = 1, \dots, M$, *be a set of data points, and let* $\lambda > 0$ *be some regularization parameter. Then, for* $\alpha > d$ *and* $\alpha - d \notin \mathbb{N}$, *the solution set* S *of the functional optimization problem*

$$\underset{f \in \mathcal{M}^\alpha(\mathbb{R}^d)}{\arg\min} \sum_{m=1}^{M} E(y_m, P_{\mathrm{Liz},\alpha}\{f\}(\mathbf{x}_m)) + \lambda \|(-\Delta)^{\alpha/2}\{f\}\|_{\mathcal{M}} \tag{5.15}$$

is nonempty and weak-compact. It is the weak* closure of the convex hull of its extreme points, which are all of the form*

$$f_{\mathrm{Ext}} = \mathrm{P}_{\mathrm{Liz},\alpha}\left\{\sum_{k=1}^{K_0} a_k(-\Delta)^{\alpha/2}\{\delta(\cdot - \mathbf{x}_k)\}\right\} = \sum_{k=1}^{K_0} a_k(\rho_{\mathrm{Liz},\alpha}(\cdot - \mathbf{x}_k) - p_{\mathbf{x}_k}) \tag{5.16}$$

for some $K_0 \leq M$, *expansion parameters* (*weights and adaptive centers*) $(a_k, \mathbf{x}_k) \in \mathbb{R} \times \mathbb{R}^d$ *for* $k = 1, \dots, K_0$, $p_{\mathbf{x}_k} \in \mathcal{P}_{\lceil \alpha - d - 1 \rceil}(\mathbb{R}^d)$, *and the radial basis function* $\rho_{\mathrm{Liz},\alpha}\colon \mathbb{R}^d \to \mathbb{R}$ *from Sec.* 4.1.

Remark 5.4. Starting from the chosen representative, we could replace $P_{\mathrm{Liz},\alpha}\{f\}(\mathbf{x}_m)$ with $P_{\mathrm{Liz},\alpha}\{f\}(\mathbf{x}_m) + p(\mathbf{x}_m)$, where $p \in \mathcal{P}_{\lfloor \alpha - d \rfloor}(\mathbb{R}^d)$, which would result in a minimization over $L_{\infty,\alpha-d}(\mathbb{R}^d)$. Hence, we are back to a more classical setting and a similar result holds, see [38, Theorem 3]. Proving the weak*-continuity of the evaluation functional in this extended setting follows along the lines of the Lizorkin-distribution setting.

5.4. *Radon splines*

The native Banach space for interpolation with Radon splines, given some order $m \in \mathbb{N}$, is the space $\mathcal{M}_{\mathrm{Rad},m}(\mathbb{R}^d)$ introduced in Sec. 4.2. Due to the form of the function $\rho_{\mathrm{Rad},m}$, this interpolation problem is closely related to approximations with 2-layer neural networks as pointed out in [1, 23]. Since the point evaluations are in the predual, the application of Theorem 5.3, together with the explicit representation of elements in $\mathcal{M}_{\mathrm{Rad},m}(\mathbb{R}^d)$ obtained in Corollary 4.2, yields the following.

Theorem 5.4 (Minimum-Energy Radon Splines). *Let* $E\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ *be a strictly convex loss function, let* $(\mathbf{x}_i, y_i) \in \mathbb{R}^d \times \mathbb{R}, i = 1, \ldots, M,$ *be a set of data points, and let* $\lambda > 0$ *be some regularization parameter. For* $m \in \mathbb{N}, m \geq 2$, *the solution set* S *of the functional optimization problem*

$$\underset{f \in \mathcal{M}_{\mathrm{Rad},m}(\mathbb{R}^d)}{\arg\min} \sum_{i=1}^{M} E(y_i, P_{\mathrm{Rad},m}\{f\}(\mathbf{x}_i)) + \lambda \|\partial_t^m \mathrm{K}_{\mathrm{rad}} \mathrm{R}\{f\}\|_{\mathcal{M}_m} \tag{5.17}$$

is nonempty and weak-compact. It is the weak* closure of the convex hull of its extreme points, which are all of the form*

$$f_{\mathrm{Ext}} = \mathrm{P}_{\mathrm{Rad},m}\left\{\sum_{k=1}^{K_0} a_k \mathrm{R}^* \partial_t^{-m}\{\delta(\cdot - \mathbf{x}_k)\}\right\} = \sum_{k=1}^{K_0} a_k(\rho_{\mathrm{Rad},m}(\langle \boldsymbol{\xi}_k, \cdot\rangle - t_k) - p_{t_k,\boldsymbol{\xi}_k}) \tag{5.18}$$

for some $K_0 \leq M$, *expansion parameters* (*weights and adaptive centers*) $(a_k, t_k, \boldsymbol{\xi}_k) \in \mathbb{R} \times \mathbb{R} \times \mathbb{S}^{d-1}$ *for* $k = 1, \ldots, K_0$, $p_{t_k,\mathbf{w}_k} \in \mathcal{P}_{m-1}(\mathbb{R}^d)$, *and the Radon radial-basis function* $\rho_{\mathrm{Rad},m}\colon \mathbb{R} \to \mathbb{R}$ *defined by* (4.26).

Remark 5.5. Starting from the chosen representative, we can also add the minimization over $\mathcal{P}_{m-1}(\mathbb{R}^d)$ to the problem and replace $\mathrm{P}_{\mathrm{Rad},m}\{f\}(\mathbf{x}_m)$ with $\mathrm{P}_{\mathrm{Rad},m}\{f\}(\mathbf{x}_m) + p(\mathbf{x}_m)$, where $p \in \mathcal{P}_{m-1}(\mathbb{R}^d)$, which results in a minimization over $L_{\infty,m-1}(\mathbb{R}^d)$. As this rules out the dependence on the representation operator $P_{\mathrm{Rad},m}$, we are back in a classical setting and a similar result holds (with $K_0 \leq M - m$), see [38, Theorem 3]. Further, we can also evaluate $\|\partial_t^m \mathrm{K}_{\mathrm{rad}} \mathrm{R}\{f\}\|_{\mathcal{M}_m}$ in the sense of $\mathcal{S}'(\mathbb{R}^d)$ since $\mathcal{P}_{m-1}(\mathbb{R}^d) \subset \ker \partial_t^m \mathrm{K}_{\mathrm{rad}} \mathrm{R}$. Compared to previous results in the literature [1, 23], this leads to a stronger characterization of the solution set S together with a nice and elegant proof.

6. Conclusions

We have shown that continuous projections onto the Lizorkin space cannot exist. Therefore, we had to resort to projection-free approaches to find representatives of Lizorkin distributions. Using the property that the space is dense in $C_0(\mathbb{R}^d)$, we have established a framework for finding representatives of distributions that lie in certain Banach subspaces. To do so, we only require representations of the related Green functions with sufficient regularity. Based on the obtained representation operator, we have introduced a powerful variational framework for the study of a wide class of inverse problems. In particular, this enabled us to strengthen results obtained in prior works. As future work, we want to apply our framework to the study of other subspaces and related variational models.

Acknowledgments

The research leading to these results has received funding from the European Research Council (ERC) under European Unions Horizon 2020 (H2020), Grant Agreement - Project No 101020573 FunLearn. Further, the authors want to thank Joachim Krieger and Marc Troyanov for fruitful discussions on the topic and in particular, Joachim Krieger for providing us with a proof of the nonexistence of projections.

Appendix A. Fundamental Solutions of the Fractional Laplacian

Given $\alpha \in \mathbb{R}$ and $d \in \mathbb{N}$ with $\alpha > d$ and $\alpha - d \notin \mathbb{N}$, we want to provide an estimate of the asymptotic behavior of $f_{\alpha,d}\colon \mathbb{R}^d \to \mathbb{R}$ with $f_{\alpha,d}(\mathbf{x}) = \|\mathbf{x}\|^{\alpha-d}$ and of its derivatives. For this purpose, we need the following lemma.

Lemma A.1. *For any* $\mathbf{k} \in \mathbb{N}^d$, *it holds that* $\partial^{\mathbf{k}}\|\cdot\| = p_{\mathbf{k}}/\|\cdot\|^{-1+2|\mathbf{k}|}$ *for some polynomial* $p \in \mathcal{P}(\mathbb{R}^d)$ *of order at most* $|\mathbf{k}|$.

Proof. We proceed by induction. For $\mathbf{k} = \mathbf{0}$ the result is obviously true. Assume that the claim holds for any $\mathbf{k} \in \mathbb{N}^d$ with $|\mathbf{k}| \leq n$ and let $\mathbf{k} \in \mathbb{N}^d$ with $|\mathbf{k}| = n+1$. For simplicity of notation, we assume that the derivative with respect to $\mathbf{x}_1$ is included and define $\tilde{\mathbf{k}} = \mathbf{k} - \mathbf{e}_1$. The induction assumption implies that

$$\partial^{\mathbf{k}}\|\mathbf{x}\| = \partial_{\mathbf{x}_k}\partial^{\tilde{\mathbf{k}}}\|\mathbf{x}\| = \partial_{\mathbf{x}_k}\frac{p_{\tilde{\mathbf{k}}}(\mathbf{x})}{\|\mathbf{x}\|^{-1+2|\tilde{\mathbf{k}}|}} = \frac{\partial_{\mathbf{x}_k}p_{\tilde{\mathbf{k}}}(\mathbf{x})\|\mathbf{x}\|^{-1+2|\tilde{\mathbf{k}}|} - p_{\tilde{\mathbf{k}}}(\mathbf{x})\mathbf{x}_k\|\mathbf{x}\|^{-3+2|\tilde{\mathbf{k}}|}}{\|\mathbf{x}\|^{-2+4|\tilde{\mathbf{k}}|}}$$

$$= \frac{p_{\mathbf{k}}(\mathbf{x})\|\mathbf{x}\|^{-3+2|\tilde{\mathbf{k}}|}}{\|\mathbf{x}\|^{-2+4|\tilde{\mathbf{k}}|}} = \frac{p_{\mathbf{k}}(\mathbf{x})}{\|\mathbf{x}\|^{-1+2|\mathbf{k}|}}, \tag{A.1}$$

which concludes the proof. □

Lemma A.1 is going to let us prove the actual result.

Proposition A.1. *For any* $\mathbf{k} \in \mathbb{N}^d$ *with* $|\mathbf{k}| \leq \lceil \alpha - d \rceil$ *and* $x \neq 0$, *it holds that* $|\partial^{\mathbf{k}} f_{\alpha,d}(\mathbf{x})| \leq C\|\mathbf{x}\|^{\alpha-d-|\mathbf{k}|}$.

Proof. We proceed inductively over $\lceil \alpha - d \rceil$. For $\lceil \alpha - d \rceil = 1$, Lemma A.1 implies that

$$|\partial_{\mathbf{x}_k} f_{\alpha,d}(\mathbf{x})| = |(\alpha - d - 1) f_{\alpha-1,d}(\mathbf{x})\partial_{\mathbf{x}_k}\|\mathbf{x}\|| \leq C\|\mathbf{x}\|^{\alpha-d-1}. \tag{A.2}$$

If $k = 0$, there is nothing to show. Assume now that the results holds for $\lceil \alpha - d \rceil = n$ and let α, d be such that $\lceil \alpha - d \rceil = n+1$. Like in the proof of Lemma A.1, we assume again that the derivative with respect to $\mathbf{x}_1$ is included and define $\tilde{\mathbf{k}} = \mathbf{k} - \mathbf{e}_1$. Then, using the Leibniz rule, we provide the estimate

$$\begin{aligned} |\partial^{\mathbf{k}} f_{\alpha,d}(\mathbf{x})| &= |\partial^{\tilde{\mathbf{k}}} \partial_{\mathbf{x}_1} f_{\alpha,d}(\mathbf{x})| \\ &\leq C \big|\partial^{\tilde{\mathbf{k}}}(f_{\alpha-1,d}(\mathbf{x}) \partial_{\mathbf{x}_1} \|\mathbf{x}\|)\big| \leq C \sum_{\mathbf{i}<\tilde{\mathbf{k}}} \big|\partial^{\tilde{\mathbf{k}}-\mathbf{i}} f_{\alpha-1,d}(\mathbf{x}) \partial^{\mathbf{i}+\mathbf{e}_1} \|\mathbf{x}\|\big| \\ &\leq C \sum_{\mathbf{i}<\tilde{\mathbf{k}}} \|\mathbf{x}\|^{\alpha-1-d-|\tilde{\mathbf{k}}-\mathbf{i}|} \|\mathbf{x}\|^{-|\mathbf{i}|} \leq C \|\mathbf{x}\|^{\alpha-d-|\mathbf{k}|}, \end{aligned} \tag{A.3}$$

which concludes the proof. □

Appendix B. Radon Transform

Here, we recall some important properties of the Radon transform, for which an extensive overview is given in [17]. The Radon transform is first described for Lizorkin functions and then extended to distributions by duality.

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

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

Its adjoint is the back-projection R^*, whose action on $g\colon \mathbb{R} \times \mathbb{S}^{d-1} \to \mathbb{R}$ is defined as

$$\mathrm{R}^*\{g\}(\mathbf{x}) = \int_{\mathbb{S}^{d-1}} g(\underbrace{\boldsymbol{\xi}^{\mathsf{T}} \mathbf{x}}_{t}, \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}, \quad \mathbf{x} \in \mathbb{R}^d. \tag{B.2}$$

Given the Fourier transform $\widehat{f} := \mathcal{F}\{f\}$ of $f \in L_1(\mathbb{R}^d)$, we can calculate $\mathrm{R}\{f\}(\cdot, \boldsymbol{\xi}_0)$ at given $\boldsymbol{\xi}_0 \in \mathbb{S}^{d-1}$ through the relation

$$\mathrm{R}\{f\}(t, \boldsymbol{\xi}_0) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{f}(\omega \boldsymbol{\xi}_0) \mathrm{e}^{\mathrm{j}\omega t} \mathrm{d}\omega = \mathcal{F}^{-1}\{\widehat{f}(\cdot \boldsymbol{\xi}_0)\}(t), \tag{B.3}$$

a property that is referred to as the *Fourier-slice theorem.* The key property for analysis purposes is that the Radon transform is continuous and invertible if the spaces are chosen properly, see [16, 17, 21] for details.

Theorem B.1 (Continuity and Invertibility of the Radon Transform on $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$). *The Radon operators* $\mathrm{R}\colon \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d) \to \mathcal{S}_{\mathrm{Liz},0}(\mathbb{R} \times \mathbb{S}^{d-1})$ *and* $\mathrm{R}^*\colon \mathcal{S}_{\mathrm{Liz},0}(\mathbb{R} \times \mathbb{S}^{d-1}) \to \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ *are bijective and continuous. Moreover*, $\mathrm{R}^* \mathrm{K}_{\mathrm{rad}} \mathrm{R} = \mathrm{K} \mathrm{R}^* \mathrm{R} = \mathrm{R}^* \mathrm{R} \mathrm{K} = \mathrm{Id}$ *on* $\mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ *and* $\mathrm{K}_{\mathrm{rad}} \mathrm{R} \mathrm{R}^* = \mathrm{Id}$ *on* $\mathcal{S}_{\mathrm{Liz},0}(\mathbb{R} \times \mathbb{S}^{d-1})$, *where* $\mathrm{K} = (\mathrm{R}^* \mathrm{R})^{-1} = c_d(-\Delta)^{(d-1)/2}$ *with* $c_d = (2(2\pi)^{d-1})^{-1}$ *is the so-called "filtering" operator and where* $\mathrm{K}_{\mathrm{rad}}$ *is an one-dimensional radial counterpart that acts along the Radon-domain variable* t*. These filtering operators are characterized by their frequency response* $\widehat{K}(\boldsymbol{\omega}) = c_d \|\boldsymbol{\omega}\|^{d-1}$ *and* $\widehat{K}_{\mathrm{rad}}(\omega) = c_d |\omega|^{d-1}$.

As evidenced in (2.8), the impulse response of the filtering operator K in Theorem B.1 is proportional to $k_{-d+1,d}$, which tells us that it asymptotically decays like

$1/\|\mathbf{x}\|^{2d-1}$ when d is even, or is a power of the Laplacian (local operator) otherwise. Further, we note that Theorem B.1 implies that R is actually a homeomorphism.

Distributional Extension This framework is extended to distributions by duality.

Definition B.1. The distribution $g = \mathrm{R}\{f\} \in \mathcal{S}'_{\mathrm{Liz},0}(\mathbb{R} \times \mathbb{S}^{d-1})$ is the Radon transform of $f \in \mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$ if

$$\forall\, \phi \in \mathcal{S}_{\mathrm{Liz},0}(\mathbb{R} \times \mathbb{S}^{d-1}) : \langle g, \phi\rangle_{\mathrm{Rad}} = \langle f, \mathrm{R}^*\{\phi\}\rangle. \tag{B.4}$$

Likewise, $\tilde{g} = \mathrm{KR}\{f\} \in \mathcal{S}'_{\mathrm{Liz},0}(\mathbb{R} \times \mathbb{S}^{d-1})$ is the filtered projection of $f \in \mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$ if

$$\forall\, \phi \in \mathcal{S}_{\mathrm{Liz},0}(\mathbb{R} \times \mathbb{S}^{d-1}) : \langle \tilde{g}, \phi\rangle_{\mathrm{Rad}} = \langle f, \mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\{\phi\}\rangle. \tag{B.5}$$

Finally, the backprojection $f = \mathrm{R}^*\{g\} \in \mathcal{S}'_{Liz}(\mathbb{R}^d)$ of $g \in \mathcal{S}'_{\mathrm{Liz},0}(\mathbb{R} \times \mathbb{S}^{d-1})$ is defined via

$$\forall\, \varphi \in \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d) : \langle \mathrm{R}^*\{g\}, \varphi\rangle = \langle g, \mathrm{R}\{\varphi\}\rangle_{\mathrm{Rad}}. \tag{B.6}$$

Due to duality, the distributional extension of the Radon transform inherits most of the properties of the "classical" operator defined by (B.1).

Theorem B.2 (Invertibility of the Radon Transform on $\mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$). *It holds that* $\mathrm{R}^*\mathrm{K}_{\mathrm{rad}}\mathrm{R} = \mathrm{KR}^*\mathrm{R} = \mathrm{Id}$ *on* $\mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$. *Hence, the "filtered-projection" operator* $\mathrm{K}_{\mathrm{rad}}\mathrm{R}\colon \mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d) \to \mathcal{S}'_{\mathrm{Liz},0}(\mathbb{R}^d)$ *is a homeomorphism with inverse* $\mathrm{R}^*\colon \mathcal{S}'_{\mathrm{Liz},0}(\mathbb{R}^d) \to \mathcal{S}'_{\mathrm{Liz}}(\mathbb{R}^d)$.

The Fourier-slice theorem expressed by (B.3) yields a unique (Fourier-based) characterization of $\mathrm{R}\{f\}$. It remains valid for tempered distributions whose generalized Fourier transforms can be identified as continuous functions of $\boldsymbol{\omega}$. It is especially helpful when the underlying function or distribution is isotropic.

An isotropic function $\rho_{\mathrm{iso}}\colon \mathbb{R}^d \to \mathbb{R}$ is characterized by its radial profile $\rho\colon \mathbb{R}_{\geq 0} \to \mathbb{R}$, so that $\rho_{\mathrm{iso}}(\mathbf{x}) = \rho(\|\boldsymbol{x}\|)$. The frequency-domain counterpart of this characterization is $\widehat{\rho}_{\mathrm{iso}}(\boldsymbol{\omega}) = \widehat{\rho}_{\mathrm{rad}}(\|\boldsymbol{\omega}\|)$ with radial frequency profile

$$\widehat{\rho}_{\mathrm{rad}}(\omega) = \frac{(2\pi)^{d/2}}{|\omega|^{d/2-1}} \int_0^{+\infty} \rho(t) t^{d/2-1} J_{d/2-1}(\omega t) t \mathrm{d}t, \tag{B.7}$$

where J_ν is the Bessel function of the first kind of order ν. In Proposition B.1, we characterize isotropic Lizorkin functions.

Proposition B.1. *Let* $\varphi_{\mathrm{iso}} \in \mathcal{S}(\mathbb{R}^d)$ *be an isotropic test function. Then,* $\varphi_{\mathrm{iso}} \in \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d)$ *if and only if* $\varphi_{\mathrm{rad}}(t) = \mathrm{R}\{\varphi_{\mathrm{iso}}\}(t, \boldsymbol{\xi}) \in \mathcal{S}_{\mathrm{Liz}}(\mathbb{R})$.

Proof. Since φ_{iso} is isotropic, for any $\mathbf{k} \in \mathbb{N}^d$, we have that

$$c_{\mathbf{k}} = \langle \mathbf{x}^{\mathbf{k}}, \varphi_{\mathrm{iso}}\rangle = \mathrm{j}^{|\mathbf{k}|}\partial^{\mathbf{k}}\widehat{\varphi}_{\mathrm{iso}}(\mathbf{0}) = \mathrm{j}^k \mathrm{D}^k\{\widehat{\varphi}_{\mathrm{rad}}\}(0) = c_k \quad \text{with } k = |\mathbf{k}|. \tag{B.8}$$

The last equality also implies that $c_k = \int_{\mathbb{R}} \varphi_{\mathrm{rad}}(t) t^k \mathrm{d}t$, where $\varphi_{\mathrm{rad}} = \mathcal{F}^{-1}\{\widehat{\varphi}_{\mathrm{rad}}\}(\cdot, \boldsymbol{\xi}) = \mathrm{R}\{\varphi\}(\cdot, \boldsymbol{\xi})$ is the radial profile (by the Fourier-slice theorem). This shows that, indeed,

$$\varphi_{\mathrm{rad}} \in \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}) \Leftrightarrow \varphi_{\mathrm{iso}} \in \mathcal{S}_{\mathrm{Liz}}(\mathbb{R}^d). \tag{B.9}$$

□

Finally, we provide a result on how to compute the Radon transform of isotropic Lizorkin distributions.

Proposition B.2 (Radon Transform of Isotropic Distributions). *Let $\rho_{\rm iso}$ be an isotropic distribution whose radial frequency profile is $\widehat{\rho}_{\rm rad}(\omega)$. Then,*

$$\mathrm{R}\{\rho_{\rm iso}(\cdot - \mathbf{x}_0)\}(t, \boldsymbol{\xi}) = \rho_{\rm rad}(t - \boldsymbol{\xi}^{\mathsf{T}}\mathbf{x}_0), \tag{B.10}$$

$$\mathrm{K}_{\rm rad}\mathrm{R}\{\rho_{\rm iso}(\cdot - \mathbf{x}_0)\}(t, \boldsymbol{\xi}\} = \tilde{\rho}_{\rm rad}(t - \boldsymbol{\xi}^{\mathsf{T}}\mathbf{x}_0), \tag{B.11}$$

$$\mathrm{R}\{\partial^{\mathbf{m}}\rho_{\rm iso}\}(t, \boldsymbol{\xi}) = \boldsymbol{\xi}^{\mathbf{m}}\mathrm{D}^{|\mathbf{m}|}\{\rho_{\rm rad}\}(t) \tag{B.12}$$

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

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

$$\widehat{\partial^{\mathbf{m}}\rho_{\rm iso}}(\omega\boldsymbol{\xi}) = (\mathrm{j}\omega\boldsymbol{\xi})^{\mathbf{m}}\widehat{\rho}_{\rm rad}(\omega) = \boldsymbol{\xi}^{\mathbf{m}}(\mathrm{j}\omega)^{|\mathbf{m}|}\widehat{\rho}_{\rm rad}(\omega), \tag{B.13}$$

which, upon taking the inverse 1D Fourier transform, yields (B.12). □

Appendix C. Extreme Points

First, we recall the definition of extreme points.

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

Proposition C.1 (Isometric Projections and Extreme Points). *Let $\mathcal{U}$ be a closed subspace of the Banach space $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ with some corresponding continuous projection $\mathrm{Proj}_{\mathcal{U}} \colon \mathcal{X} \to \mathcal{U}$. Then, the following hold:*

(1) *The unit ball in the Banach space $\mathcal{U} = \mathrm{Proj}_{\mathcal{U}}(\mathcal{X})$ satisfies*

$$B_{\mathcal{U}}(1) \subseteq \mathrm{Proj}_{\mathcal{U}}(B_{\mathcal{X}}(1)) \subseteq B_{\mathcal{U}}(\|\mathrm{Proj}_{\mathcal{U}}\|), \tag{C.1}$$

where $B_{\mathcal{U}}(r) = \{x \in \mathcal{U} : \|u\|_{\mathcal{U}} \le r\}$ and $\|\mathrm{Proj}_{\mathcal{U}}\|$ is the norm of the underlying projector. Consequently, $B_{\mathcal{U}}(1) = \mathrm{Proj}_{\mathcal{U}}(B_{\mathcal{X}}(1))$ if and only if $\|\mathrm{Proj}_{\mathcal{U}}\| = 1$.

(2) *Let $\tilde{E} = \{\mathrm{Proj}_{\mathcal{U}}\{e\} : e \in \mathrm{Ext}(B_{\mathcal{X}}(1))\}\backslash\{0\}$. If $\|\mathrm{Proj}_{\mathcal{U}}\| = 1$ and all $\tilde{e} \in \tilde{E}$ satisfy $\|\tilde{e}\|_{\mathcal{X}} = 1$, then $B_{\mathcal{U}}(1)$ is the closed convex hull of $\tilde{E}$ so that $\mathrm{Ext}(B_{\mathcal{U}}(1)) \subseteq \tilde{E}$.*

Proof. For the first statement, note that the unit ball in $\mathcal{U}$ is $B_{\mathcal{U}}(1) = B_{\mathcal{X}}(1) \cap \mathcal{U}$. In particular, $u = \mathrm{Proj}_{\mathcal{U}}\{u\}$ and $\|u\|_{\mathcal{X}} \le 1$ for any $u \in B_{\mathcal{U}}(1)$, which implies that $B_{\mathcal{U}}(1) \subseteq \mathrm{Proj}_{\mathcal{U}}(B_{\mathcal{X}}(1))$. Next, we recall that the norm of $\mathrm{Proj}_{\mathcal{U}} \colon \mathcal{X} \to \mathcal{U}$ is given by

$$\|\mathrm{Proj}_{\mathcal{U}}\| = \sup_{x \in \mathcal{X}\backslash\{0\}} \frac{\|\mathrm{Proj}_{\mathcal{U}}\{x\}\|_{\mathcal{X}}}{\|x\|_{\mathcal{X}}}. \tag{C.2}$$

Therefore, any $x \in B_{\mathcal{X}}(1)$ satisfies $\|\mathrm{Proj}_{\mathcal{U}}\{x\}\|_{\mathcal{X}} \le \|\mathrm{Proj}_{\mathcal{U}}\|\,\|x\|_{\mathcal{X}} \le \|\mathrm{Proj}_{\mathcal{U}}\|$, which implies that $\mathrm{Proj}_{\mathcal{U}}(B_{\mathcal{X}}(1)) \subseteq B_{\mathcal{U}}(\|\mathrm{Proj}_{\mathcal{U}}\|)$.

The Krein–Milman theorem ensures that $B_{\mathcal{X}}(1)$ is the closed convex hull of its extreme points $e_k \in E = \mathrm{Ext}(B_{\mathcal{X}}(1))$, namely, $B_{\mathcal{X}}(1) = \mathrm{cch}E$. Due to $\|\mathrm{Proj}_{\mathcal{U}}\| = 1$, it holds that $B_{\mathcal{U}}(1) = \mathrm{Proj}_{\mathcal{U}}(B_{\mathcal{X}}(1))$. Further, as $B_{\mathcal{U}}(1)$ is convex, each $u = \mathrm{Proj}_{\mathcal{U}}(\sum_{k=1}^{K} \theta_k e_k) = \sum_{k=1}^{K} \theta_k \tilde{e}_k$ with $\theta_k \geq 0$, $\sum_{k=1}^{K} \theta_k = 1$ and $e_k \in E$ lies in $B_{\mathcal{U}}(1)$. In other words, the convex hull of the e_k maps onto the convex hull of the $\tilde{e}_k = \mathrm{Proj}_{\mathcal{U}} e_k$ with $\mathrm{ch}\{\tilde{e}_k\} \subseteq B_{\mathcal{U}}(1)$. Since $\mathrm{Proj}_{\mathcal{U}}$ is continuous and $B_{\mathcal{X}}(1)$ is closed, the argument carries over to limits as well. Hence, it holds that $B_{\mathcal{U}}(1) = \mathrm{Proj}_{\mathcal{U}}(\mathrm{cch}E) = \mathrm{cch}(\mathrm{Proj}_{\mathcal{U}}(E)) = \mathrm{cch}(\tilde{E})$. □

Example C.1. The projector $\mathrm{Proj}_{\mathrm{even}} \colon \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}_{\mathrm{even}}(\mathbb{R}^d)$ onto the even Schwartz functions is given by

$$\mathrm{Proj}_{\mathrm{even}}\{f\}(\mathbf{x}) = \frac{f(\mathbf{x}) + f(-\mathbf{x})}{2}. \tag{C.3}$$

By duality, we define $\mathrm{Proj}_{\mathrm{even}} \colon \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'_{\mathrm{even}}(\mathbb{R}^d)$. The extreme points of $\mathcal{M}(\mathbb{R}^d)$ are $(\delta(\cdot - \tau))_{\tau \in \mathbb{R}^d}$. Since

$$\|\mathrm{Proj}_{\mathrm{even}}\{\delta(\cdot - \tau)\}\|_{\mathcal{M}} = \left\| \frac{1}{2}\delta(\cdot + \tau) + \frac{1}{2}\delta(\cdot - \tau) \right\|_{\mathcal{M}} = 1 \tag{C.4}$$

for all $\tau \in \mathbb{R}^d$, the extreme points of $\mathcal{M}_{\mathrm{even}}(\mathbb{R}^d)$ are of the form $\frac{1}{2}\delta(\cdot + \tau) + \frac{1}{2}\delta(\cdot - \tau)$ with $\tau \in \mathbb{R}^d$.

References

[1] F. Bartolucci, E. De Vito, L. Rosasco and S. Vigogna, Understanding neural networks with reproducing kernel Banach spaces, preprint (2021), arXiv:2109.09710.

[2] F. Bartolucci, S. Pilipović and N. Teofanov, The Shearlet transform and Lizorkin spaces, in *Landscapes of Time-Frequency Analysis*, Applied and Numerical Harmonic Analysis (Springer, Cham, 2020), pp. 43–62.

[3] P. Bohra, J. Campos, H. Gupta, S. Aziznejad and M. Unser, Learning activation functions in deep (spline) neural networks, *IEEE Open J. Signal Process.* **1** (2020) 295–309.

[4] K. Bredies and M. Carioni, Sparsity of solutions for variational inverse problems with finite-dimensional data, *Calc. Var. Partial Differ. Equ.* **59**(1) (2020) 14.

[5] L. Bungert, T. Roith, D. Tenbrinck and M. Burger, A Bregman learning framework for sparse neural networks, *J. Mach. Learn. Res.* **23**(192) (2022) 1–43.

[6] P. L. Combettes, S. Salzo and S. Villa, Regularized learning schemes in feature Banach spaces, *Anal. Appl. (Singap.)* **16**(1) (2018) 1–54.

[7] T. Debarre, J. Fageot, H. Gupta and M. Unser, B-spline-based exact discretization of continuous-domain inverse problems with generalized TV regularization, *IEEE Trans. Inform. Theory* **65**(7) (2019) 4457–4470.

[8] Q. Denoyelle, V. Duval, G. Peyré and E. Soubies, The sliding Frank–Wolfe algorithm and its application to super-resolution microscopy, *Inv. Prob.* **36**(1) (2020) 014001.

[9] J. Fageot and M. Simeoni, TV-based reconstruction of periodic functions, *Inverse Probl.* **36**(11) (2020) 115015.

[10] J. Fageot, M. Unser and J. Ward, Beyond Wiener's lemma: Nuclear convolution algebras and the inversion of digital filters, *J. Fourier Anal. Appl.* **25**(4) (2019) 2037–2063.

[11] M. Fey, J. E. Lenssen, F. Weichert and H. Müller, SplineCNN: Fast geometric deep learning with continuous B-spline kernels, in *IEEE Conf. Computer Vision and Pattern Recognition* (IEEE, 2018), pp. 869–877.

[12] A. Flinth, F. de Gournay and P. Weiss, On the linear convergence rates of exchange and continuous methods for total variation minimization, *Math. Program.* **190**(1–2, Ser. A) (2021) 221–257.

[13] I. M. Gelfand and G. Shilov, *Generalized Functions. Vol. 1. Properties and Operations* (Academic Press, New York, USA, 1964).

[14] A. Grothendieck, Produits tensoriels topologiques et espaces nucléaires, *Mem. Amer. Math. Soc.* **16** (1955) 1–336.

[15] H. Gupta, J. Fageot and M. Unser, Continuous-domain solutions of linear inverse problems with Tikhonov versus generalized TV regularization, *IEEE Trans. Signal Process.* **66**(17) (2018) 4670–4684.

[16] S. Helgason, The Radon transform on Euclidean spaces, compact two-point homogeneous spaces and Grassmann manifolds, *Acta Math.* **113** (1965) 153–180.

[17] S. Helgason, *The Radon Transform*, Progress in Mathematics, Vol. 5 (Birkhäuser, Boston, 1999).

[18] S. Kostadinova, S. Pilipović, K. Saneva and J. Vindas, The ridgelet transform of distributions, *Integral Transforms Spec. Funct.* **25**(5) (2014) 344–358.

[19] R. Lin, H. Zhang and J. Zhang, On reproducing kernel Banach spaces: Generic definitions and unified framework of constructions, preprint (2019), arXiv:1901.01002.

[20] P. I. Lizorkin, Generalized Liouville differentiation and the functional spaces $L_p^r(E_n)$. Imbedding theorems, *Mat. Sb. (N.S.)* **60**(102) (1963) 325–353.

[21] D. Ludwig, The Radon transform on Euclidean space, *Commun. Pure Appl. Math.* **19**(1) (1966) 49–81.

[22] G. Ongie, R. Willett, D. Soudry and N. Srebro, A function space view of bounded norm infinite width ReLU nets: The multivariate case, in *8th International Conference on Learning Representations*, (*ICLR*), Addis Ababa, Ethiopia, 26–30 April 2020.

[23] R. Parhi and R. D. Nowak, Banach space representer theorems for neural networks and ridge splines, *J. Mach. Learn. Res.* **22**(41) (2021) 1–40.

[24] R. Parhi and R. D. Nowak, What kinds of functions do deep neural networks learn? Insights from variational spline theory, preprint (2021), arXiv:2105.03361.

[25] M. Reed and B. Simon, *Methods of Modern Mathematical Physics. I*, 2nd edn. (Academic Press, New York, 1980).

[26] B. Rubin, *Fractional Integrals and Potentials*, Pitman Monographs and Surveys in Pure and Applied Mathematics, Vol. 82 (Longman, Harlow, 1996).

[27] S. Samko, Denseness of the spaces Φ_V of Lizorkin type in the mixed $L^{\overline{p}}(\mathbb{R}^n)$-spaces, *Studia Math.* **113**(3) (1995) 199–210.

[28] S. G. Samko, *Hypersingular Integrals and their Applications*, Analytical Methods and Special Functions, Vol. 5 (Taylor & Francis Group, London, 2002).

[29] S. G. Samko, A. A. Kilbas and O. I. Marichev, *Fractional Integrals and Derivatives: Theory and Applications* (Gordon and Breach Science Publishers, 1993).

[30] P. Savarese, I. Evron, D. Soudry and N. Srebro, How do infinite width bounded norm networks look in function space? in *Proc. Thirty-Second Conf. Learning Theory*, Proceedings of Machine Learning Research, Vol. 99 (PMLR, 2019), pp. 2667–2690.

[31] I. J. Schoenberg, Contributions to the problem of approximation of equidistant data by analytic functions, *Q. Appl. Math.* **4** (1946) 45–99.

[32] L. Schwartz, *Théorie des Distributions*, Publications de l'Institut de Mathématique de l'Université de Strasbourg, IX-X (Hermann, Paris, 1966).

[33] S. Sonoda and N. Murata, Neural network with unbounded activation functions is universal approximator, *Appl. Comput. Harmon. Anal.* **43**(2) (2017) 233–268.

[34] P. R. Stinga, User's guide to the fractional Laplacian and the method of semigroups, in *Handbook of Fractional Calculus with Applications*, Vol. 2 (De Gruyter, Berlin, 2019), pp. 235–265.

[35] F. Trèves, *Topological Vector Spaces, Distributions and Kernels* (Dover Publications, New York, 2006).

[36] M. Troyanov, On the Hodge decomposition in $\mathbb{R}^n$, *Mosc. Math. J.* **9**(4) (2009) 899–926.

[37] M. Unser, A unifying representer theorem for inverse problems and machine learning, *Found. Comput. Math.* **21**(4) (2021) 941–960.

[38] M. Unser and S. Aziznejad, Convex optimization in sums of Banach spaces, *Appl. Numer. Harmon. Anal.* **56** (2022) 1–25.

[39] M. Unser and T. Blu, Fractional splines and wavelets, *SIAM Rev.* **42**(1) (2000) 43–67.

[40] M. Unser, J. Fageot and J. P. Ward, Splines are universal solutions of linear inverse problems with generalized TV regularization, *SIAM Rev.* **59**(4) (2017) 769–793.

[41] W. Yuan, W. Sickel and D. Yang, *Morrey and Campanato Meet Besov, Lizorkin and Triebel*, Lecture Notes in Mathematics, Vol. 2005 (Springer-Verlag, Berlin, 2010).

[42] H. Zhang, Y. Xu and J. Zhang, Reproducing kernel Banach spaces for machine learning, *J. Mach. Learn. Res.* **10** (2009) 2741–2775.