SPLINES ARE UNIVERSAL SOLUTIONS OF LINEAR INVERSE PROBLEMS WITH GENERALIZED-TV REGULARIZATION

MICHAEL UNSER*, JULIEN FAGEOT*, AND JOHN PAUL WARD*†

Abstract. Splines come in a variety of flavors that can be characterized in terms of some differential operator L. The simplest piecewise-constant model corresponds to the derivative operator. Likewise, one can extend the traditional notion of total variation by considering more general operators than the derivative. This results in the definition of a generalized total variation semi-norm and of its corresponding native space, which is further identified as the direct sum of two Banach spaces. We then prove that the minimization of the generalized total variation (gTV), subject to some arbitrary (convex) consistency constraints on the linear measurements of the signal, admits nonuniform L-spline solutions with fewer knots than the number of measurements. This shows that nonuniform splines are universal solutions of continuous-domain linear inverse problems with LASSO, L_1, or total-variation-like regularization constraints. Remarkably, the type of spline is fully determined by the choice of L and does not depend on the actual nature of the measurements.

Key words. Sparsity, total variation, splines, inverse problems, compressed sensing

AMS subject classifications. 41A15, 47A52, 94A20, 46E27, 46N20, 47F05, 34A08, 26A33

1. Introduction. Imposing sparsity constraints is a powerful paradigm for solving ill-posed inverse problems and/or for reconstructing signals at super-resolution [6]. This is usually achieved by formulating the task as an optimization problem that includes some form of ℓ_1 regularization [49]. The concept is central to the theory of compressed sensing (CS) [9, 19] and is currently driving the development of a new generation of algorithms for the reconstruction of biomedical images [36]. The primary factors that have contributed to making sparsity a remarkably popular research topic during the past decade are as follows:

- the possibility of recovering the signal from few measurements (CS) with a theoretical guarantee of perfect recovery under strict conditions [7, 10, 19];
- the availability of fast iterative solvers for this class of problems [4, 14, 26, 39];
- the increasing evidence of the superiority of the sparsity-promoting schemes over the classical linear reconstruction (including the Tikhonov ℓ_2 regularization) in a variety of imaging modalities [36].

The approach developed in this paper is also driven by the idea of sparsity. However, it deviates from the standard paradigm because the recovery problem is formulated in the continuous domain under the practical constraint of a finite number of linear measurements. The ill-posedness of the problem is then dealt with by searching for a solution that is consistent with the measurements and that minimizes a generalized version of the total-variation (TV) semi-norm—the continuous-domain counterpart of ℓ_1 regularization. Our major finding (Theorem 1) is that the extremal points of this kind of recovery problem are nonuniform splines whose type is matched to the regularization operator L. The powerful aspect is that the result holds in full generality, as long as the problem remains convex. The only constraint is that the linear inverse problem should be well-posed over the (very small) null space of the regularization operator, which is the minimal requirement for any valid regularization scheme. In

*Biomedical Imaging Group, École polytechnique fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland,

†Department of Mathematics, University of Central Florida, Orlando, FL, USA. Present address: Department of Mathematics, North Carolina A&T State University, Greensboro, NC, USA.

particular, Theorem 1 gives a theoretical explanation of the well-documented observation that total variation regularization—the simplest case of the present theory with L = D (derivative operator)—tends to produce piecewise-constant solutions [11, 41]. Recognizing the intimate connection between linear inverse problems and splines is also helpful for discretization purposes because it provides us with a parametric representation of the solution that is controlled by the regularization operator L. In that respect, our representer theorems extend some older results on spline interpolation with minimum L_1-norms, including the adaptive regression splines of Mammen and van de Geer [38] and the functional analytic characterization of Fisher and Jerome [27]. There is a connection as well with the work of Steidl *et al.* on splines and higher-order TV [48], although their formulation is strictly discrete and restricted to the denoising problem.

2. Linear Inverse Problems: Current Status and Motivation. Our notational convention is to use bold letters to denote ordinary vectors and matrices to distinguish them from their infinite-dimensional counterparts; that is, functions (such as s) and linear operators (such as L). Simply stated, the inverse problem is to recover a signal s from a finite set of linear measurements $\mathbf{y} = \mathbf{y}_0(s) + \mathbf{n} \in \mathbb{R}^M$ where $\mathbf{n}$ is a disturbance term that is usually assumed to be small and independent of s. In most real-world problem the unknown signal lives in the continuum so that it is appropriate to view it as an element of some Banach space $\mathcal{B}$. Then, by the assumption of linearity, there exists a set of functionals $\nu_m \in \mathcal{B}'$ (the continuous dual of $\mathcal{B}$) with $m = 1, \ldots, M$ such that the noise-free measurements are given by $\mathbf{y}_0 = \boldsymbol{\nu}(s) = (\langle \nu_1, s \rangle, \ldots, \langle \nu_M, s \rangle)$. The measurement functionals ν_m are governed by the underlying physics (forward model) and assumed to be known. Since the signal $s \in \mathcal{B}$ is an infinite-dimensional entity and the number of measurements is finite, the inverse problem is obviously ill-posed, not to mention the fact that the true measurements $\mathbf{y}$ are typically only approximate versions of $\mathbf{y}_0$ since they are corrupted by noise.

2.1. Finite-Dimensional Formulation. The standard approach for the resolution of such inverse problems is to select some finite-dimensional reconstruction space $\mathcal{V} = \text{span}\{\varphi_n\}_{n=1}^N \subset \mathcal{B}$. Based on the (simplifying) assumption that $s \in \mathcal{V}$, one then converts the original noise-free forward model into the discretized version $\mathbf{y}_0 = \mathbf{A}\mathbf{x}$, where $\mathbf{x} \in \mathbb{R}^N$ represents the expansion coefficients of s in the basis $\{\varphi_n\}_{n=1}^N$. Here, $\mathbf{A}$ is the so-called *sensing matrix* of size $(M \times N)$ whose entries are given by $[\mathbf{A}]_{m,n} = \langle \nu_m, \varphi_n \rangle$.

The basic assumption made by the theory of compressed sensing is that there exists a finite-dimensional basis (or dictionary) $\{\varphi_n\}_{n=1}^N$ that "sparsifies" the class of desired signals with the property that $\|\mathbf{x}\|_0 \leq K_0$ for some fixed K_0 which is (much) smaller than N; in other words, it should be possible to *synthesize* the signal exactly by restricting the expansion to no more than K_0 atoms in the basis $\{\varphi_n\}_{n=1}^N$ [20, 23, 40]. The signal recovery is then recast as the constrained optimization problem

$$\arg\min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \leq \epsilon^2, \tag{1}$$

where the minimization of the ℓ_1 norm promotes sparse solutions [49]. The role of the right-hand-side inequality is to encourage consistency between the noisy measurements $\mathbf{y}$ and their noise-free restitution $\mathbf{y}_0 = \mathbf{A}\mathbf{x}$. The popularity of (1) stems from the fact that the theory of CS guarantees a faithful signal recovery from $M > 2K_0$ measurements under strict conditions on $\mathbf{A}$ (*i.e.*, restricted isometry) [7, 10, 19].

Instead of basing the recovery on the synthesis formula $s = \sum_n x_n \varphi_n \in \mathcal{V}$, one can

adopt an alternative *analysis* or *regularization* point of view. To that end, one typically assumes that s is discretized in some implicit "pixel" basis with expansion coefficients $\mathbf{s} = (s_1, \ldots, s_N) \in \mathbb{R}^N$, where the s_n are the samples of the underlying signal. The corresponding system matrix (forward model) is denoted by $\mathbf{H} : \mathbb{R}^N \to \mathbb{R}^M$. Given some appropriate regularization operator $\mathbf{L} : \mathbb{R}^N \to \mathbb{R}^{N'}$, the idea then is to exploit the property that the transformed version of the signal, $\mathbf{Ls}$, is sparse. This translates into the optimization problem

$$\arg\min_{\mathbf{s}\in\mathbb{R}^N} \|\mathbf{Ls}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{Hs}\|_2^2 \leq \epsilon^2, \tag{2}$$

which is slightly more involved than (1). The two forms are equivalent only when $N' = N$ and $\mathbf{L}$ is invertible, the connection being $\mathbf{A} = \mathbf{HL}^{-1}$. For computational purposes, (2) is often converted into the equivalent unconstrained version of the problem

$$\arg\min_{\mathbf{s}\in\mathbb{R}^N} \left(\|\mathbf{y} - \mathbf{Hs}\|_2^2 + \lambda\|\mathbf{Ls}\|_1\right), \tag{3}$$

where $\lambda \in \mathbb{R}^+$ is an adjustable regularization parameter that needs to tuned such that $\|\mathbf{y}-\mathbf{Hs}\|_2^2 = \epsilon^2$. One of the preferred choices for $\mathbf{L}$ is the finite-difference operator—or the discrete version of the gradient in dimensions higher than one. This corresponds to the "total-variation" reconstruction method, which is widely used in applications [3, 11, 30, 41].

The sparsity-promoting effect of these discrete formulations and the conditions under which the expansion coefficients of the signal can be recovered are fairly well understood [28, 54]. What is less satisfactory is the intrinsic interdependence between the sparsity constraints and the choice of the appropriate reconstruction space, which makes it difficult to deduce rates of convergence and error estimates relating to the underlying continuous-domain recovery problem.

2.2. Infinite-Dimensional Formulation. Recently, Adcock and Hansen have addressed the above limitation by formulating an infinite-dimensional theory of CS [1]. The measurements are the same as before, but the unknown signal is now a function $s : \mathbb{R}^d \to \mathbb{R}$. For the purpose of illustration, we take $d = 1$ and $s \in \mathrm{BV}(\mathbb{R})$ with the property that such a function admits the (unique) expansion $s = \sum_n w_n\psi_n$ in the (properly normalized) Haar wavelet basis $\{\psi_n\}$. It is known that the condition $s \in \mathrm{BV}(\mathbb{R})$ implies the inclusion of $w = (w_n)$ in weak-$\ell_1(\mathbb{Z})$—a space that is slightly larger than $\ell_1(\mathbb{Z})$ [12]. Conversely, one can force the inclusion in $\mathrm{BV}(\mathbb{R})$ by imposing a bound on the ℓ_1 norm of these coefficients. If one further assumes that the signal is sparse in the Haar basis, one can recast the reconstruction problem as

$$\min_{w\in\ell_1(\mathbb{Z})} \|w\|_1 \quad \text{s.t.} \quad \sum_{m=1}^{M} \left|y_m - \langle \nu_m, \sum_n w_n\psi_n\rangle\right|^2 \leq \epsilon^2, \tag{4}$$

which is the infinite-dimensional counterpart of the synthesis formulation (1). The key question is to derive conditions on how to choose the ν_m to guarantee recovery of wavelet coefficients up to a certain scale. This has been done in [1, 2], which means that the issue of convergence is now reasonably well understood for the synthesis form of the recovery problem.

In our framework, $\mathcal{M}_{\mathrm{D}}(\mathbb{R})$ is the space of functions on $\mathbb{R}$ of bounded (total) variation, which is slightly larger than $\mathrm{BV}(\mathbb{R})$ because it also includes constant signals. This allows us to close the circle by enforcing a regularization on the "true" total

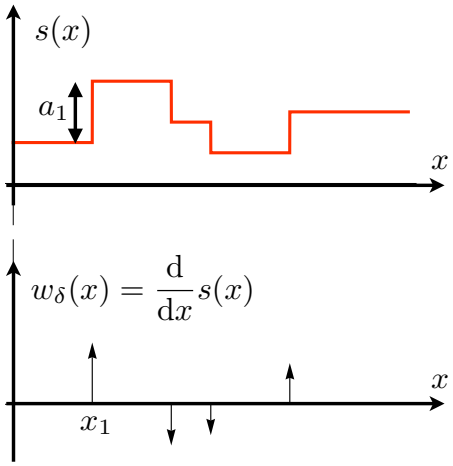

FIG. 1. *Prototypical solution of a linear inverse problem with total-variation regularization. The signal is piecewise-constant; in other words, it is a nonuniform* L*-spline with* $\mathrm{L} = \mathrm{D}$ *(derivative operator). The application of* D *uncovers the innovation* w_δ*: The Dirac impulses are located at the points of discontinuity (knots), while their height (weight) encodes the magnitude of the corresponding jump.*

variation of the solution, which is associated with the derivative operator $\mathrm{D} = \frac{\mathrm{d}}{\mathrm{d}x}$. This results in the functional optimization problem

(5)
$$s = \arg \min_{f \in \mathcal{M}_{\mathrm{D}}(\mathbb{R})} \mathrm{TV}(f) = \|\mathrm{D}f\|_{\mathcal{M}} \quad \text{s.t.} \quad \|\mathbf{y} - \boldsymbol{\nu}(f)\|_2^2 = \sum_{m=1}^{M} \left| y_m - \langle \nu_m, f \rangle \right|^2 \leq \epsilon^2,$$

which is the continuous-domain counterpart of (2). Now, the motivation for our present theory is that (5) corresponds to a special case of Theorem 1 with $\mathrm{L} = \mathrm{D}$ and the closed compact convex set $\mathcal{C} \subset \mathbb{R}^M$ being specified by the inequality on the right-hand-side of (5); that is, $\mathcal{C}_\epsilon(\mathbf{y}) = \{\mathbf{z} \in \mathbb{R}^M : \|\mathbf{y} - \mathbf{z}\|_2^2 \leq \epsilon^2\}$. The key is that the differentiation operator D is spline-admissible in the sense of Definition 1: Its causal Green's function is the Heaviside (or unit-step) function $\rho_{\mathrm{D}}(x) = \mathbb{1}_+(x)$ whose rate of growth is $n_0 = 0$, while its null space $\mathcal{N}_{\mathrm{D}} = \mathrm{span}\{p_1\}$ with $p_1(x) = 1$ is composed of all constant-valued signals. This implies that the extreme points of (5) necessarily take the form

(6)
$$s(x) = b_1 + \sum_{k=1}^{K} a_k \mathbb{1}_+(x - x_k)$$

with $K \leq M$. This corresponds to a piecewise-constant signal with jumps of size a_k at the x_k, as illustrated in Figure 1. The solution also happens to be a polynomial spline of degree 0 with knots at the x_k with the property that $\mathrm{D}\{s\} = \sum_{k=1}^{K} a_k \delta(\cdot - x_k) = w_\delta$, which is a weighted sum of shifted Dirac impulses (the innovation of the spline), as shown on the bottom of Figure 1. In view of Definition 2, the solution (6) can also be described as a nonuniform L-spline with $\mathrm{L} = \mathrm{D}$. The remarkable aspect of this result is that the parametric form (6) is universal, in the sense that it does not dependent on the measurement functionals ν_m. To the best of our knowledge,

this is the first mathematical explanation of the well-known observation that TV regularization tends to enforce piecewise-constant solutions. The other interesting point is that one can interpret the solution as the best K-term representation of the signal within an infinite-dimensional dictionary that consists of a constant signal p_1 plus a continuum of shifted Green's functions (*i.e.*, $\{\mathbb{1}_+(\cdot - \tau)\}_{\tau \in \mathbb{R}}$), making the connection with the synthesis views (1) and (4) of the problem. Also, note that the described sparsifying effect is much more dramatic than that of the finite-dimensional setting since one is collapsing a continuum (integral representation) into a discrete and finite sum.

We shall now show that the mechanism at play is very general and transposable to a much broader class of regularization operators L and data-fidelity terms, as well as for the multidimensional setting.

2.3. Road Map of the Paper. The remainder of this paper is organized as follows: After setting the notation, we present and discuss of our main representer theorem (Theorem 1) in Section 3. We also provide a refined version for the simpler interpolation scenario (Theorem 2). We then proceed with the review of primary applications in Section 4.

The mathematical tools for proving our results are developed in the second half of the paper. The first enabling component is the tight connection between splines and operators, which is the topic of Section 5. In particular, we present an operator-based method to synthesize a spline from its innovation, which requires the construction of an appropriate right-inverse operator (Theorem 4). The existence of such inverse operators is fundamental to the characterization of the native spaces associated with our generalized total-variation criterion (gTV) (Theorem 5), as we show in Section 6. The actual proof of Theorems 1 and 2 is given in Section 7. It relies on a preparatory result (generalized Fisher-Jerome theorem) that establishes the impulsive form of the solutions of some abstract minimization problem over the space $\mathcal{M}(\mathbb{R}^d)$ of bounded Borel measures.

We conclude the paper in Section 8 with a brief discussion of open issues.

3. Representer Theorems for Generalized Total Variation. Although we are considering a finite number of measurements, we are formulating the reconstruction problem in the continuous domain. This calls for a precise specification of the underlying functional setting.

3.1. Notation. The space of tempered distribution is denoted by $\mathcal{S}'(\mathbb{R}^d)$ where d gives the number of dimensions. This space is made of continuous linear functionals $\mu : \varphi \mapsto \langle \mu, \varphi \rangle$ acting on the Schwartz' space $\mathcal{S}(\mathbb{R}^d)$ of smooth and rapidly decaying test functions on $\mathbb{R}^d$ [29, 33].

We shall primarily work with the space $\mathcal{M}(\mathbb{R}^d)$ of regular, real-valued, countably additive Borel measures on $\mathbb{R}^d$, which is also known (by the Riesz-Markov theorem) to be the continuous dual of $C_0(\mathbb{R}^d)$: the Banach space of continuous functions on $\mathbb{R}^d$ that vanish at infinity equipped with the supremum norm $\|\cdot\|_\infty$ [42, Chap. 6]. Since $\mathcal{S}(\mathbb{R}^d)$ is dense in $C_0(\mathbb{R}^d)$, this allows us to define $\mathcal{M}(\mathbb{R}^d)$ as

$$\mathcal{M}(\mathbb{R}^d) = \{w \in \mathcal{S}'(\mathbb{R}^d) : \|w\|_{\mathcal{M}} = \sup_{\varphi \in \mathcal{S}(\mathbb{R}^d) : \|\varphi\|_\infty = 1} \langle w, \varphi \rangle < \infty\}, \tag{7}$$

and also to extend the space of test functions to $\varphi \in C_0(\mathbb{R}^d)$. The action of w will be denoted by $\varphi \mapsto \langle w, \varphi \rangle = \int_{\mathbb{R}^d} \varphi(\boldsymbol{x}) w(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$ where the right-hand side stands for the

Lebesgue integral of φ with respect to the underlying measure[1]. The bottom line is that $\mathcal{M}(\mathbb{R}^d)$ is the Banach space associated with the norm $\|\cdot\|_{\mathcal{M}}$ which returns the "total variation" of the measure that specifies w.

Two key observations in relation to our goal are:

1. the compatibility of the L_1 and total-variation norms with the former being stronger than the latter. Indeed, $\|f\|_{L_1(\mathbb{R}^d)} = \|f\|_{\mathcal{M}}$ for all $f \in L_1(\mathbb{R}^d)$;
2. the inclusion of Dirac impulses in $\mathcal{M}(\mathbb{R}^d)$, but not in $L_1(\mathbb{R}^d)$. Specifically, $\delta(\cdot - \boldsymbol{x}_0) \in \mathcal{M}(\mathbb{R}^d)$ for any fixed offset $\boldsymbol{x}_0 \in \mathbb{R}^d$ with $\langle \delta(\cdot - \boldsymbol{x}_0), \varphi \rangle = \varphi(\boldsymbol{x}_0)$ for all $\varphi \in C_0(\mathbb{R}^d)$.

We shall monitor the algebraic rate of growth/decay of (ordinary) functions of the variable $\boldsymbol{x} \in \mathbb{R}^d$ by verifying their inclusion in the Banach space

$$L_{\infty,\alpha_0}(\mathbb{R}^d) = \{f : \mathbb{R}^d \to \mathbb{R} \ \text{ s.t. } \ \|f\|_{\infty,\alpha_0} < +\infty\}, \tag{8}$$

where

$$\|f\|_{\infty,\alpha_0} = \operatorname*{ess\,sup}_{\boldsymbol{x}\in\mathbb{R}^d} \left(|f(\boldsymbol{x})|(1+\|\boldsymbol{x}\|)^{-\alpha_0}\right)$$

with $\alpha_0 \in \mathbb{R}$. For instance, $\boldsymbol{x}^{\boldsymbol{m}} = x_1^{m_1} \cdots x_d^{m_d} \in L_{\infty,\alpha_0}(\mathbb{R}^d)$ for $\alpha_0 \geq |\boldsymbol{m}| = m_1 + \cdots + m_d$.

A linear operator whose output is a function is represented with a roman capital letter (*e.g.*, L). The action of L on the signal s is denoted by $s \mapsto \mathrm{L}\{s\}$, or L$s$ for short. Such an operator is said to be *shift-invariant* if it commutes with the shift operator $s \mapsto s(\cdot - \boldsymbol{x}_0)$; that is, if $\mathrm{L}\{s(\cdot - \boldsymbol{x}_0)\} = \mathrm{L}\{s\}(\cdot - \boldsymbol{x}_0)$ for any admissible signal s and $\boldsymbol{x}_0 \in \mathbb{R}^d$.

3.2. Main Result on the Optimality of Splines. Since the solution is regularized, the constrained minimization is performed over some native space $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ that is tied to some admissible differential operator L, such as D, D^2 (second derivative), or Δ (Laplacian) for $d > 1$.

Definition 1 (Spline-admissible operator). *A linear operator* $\mathrm{L} : \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \to \mathcal{M}(\mathbb{R}^d)$, *where* $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \supset \mathcal{S}(\mathbb{R}^d)$ *is an appropriate subspace of* $\mathcal{S}'(\mathbb{R}^d)$, *is called* spline-admissible *if*

1. *it is shift-invariant;*
2. *there exists a function* $\rho_{\mathrm{L}} : \mathbb{R}^d \to \mathbb{R}$ *of slow growth (the Green's function of* L*) such that* $\mathrm{L}\{\rho_{\mathrm{L}}\} = \delta$, *where* δ *is the Dirac impulse. The rate of polynomial growth of* ρ_{L} *is* $n_0 = \inf\{n \in \mathbb{N} : \rho_{\mathrm{L}} \in L_{\infty,n}(\mathbb{R}^d)\}$.
3. *the (growth-restricted) null space of* L,

$$\mathcal{N}_{\mathrm{L}} = \{q \in L_{\infty,n_0}(\mathbb{R}^d) : \mathrm{L}\{q\} = 0\},$$

has the finite dimension $N_0 \geq 0$.

The native space of L, $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$, is then specified as

$$\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) = \{f \in L_{\infty,n_0}(\mathbb{R}^d) : \|\mathrm{L}f\|_{\mathcal{M}} < \infty\}. \tag{9}$$

It is largest function space for which the generalized total variation

$$\mathrm{gTV}(f) = \|\mathrm{L}f\|_{\mathcal{M}}$$

[1] The use of $w(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$ in the integral is a slight abuse of notation when the measure is not absolutely continuous with respect to the Lebesgue measure.

is well-defined under the finite-dimensional null-space constraint

$$\|\mathrm{L}f\|_{\mathcal{M}} = 0 \Leftrightarrow f \in \mathcal{N}_{\mathrm{L}}, \text{ for any } f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d).$$

This also means that gTV is only a semi-norm on $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$. However, it can be turned into a proper norm by factoring out the null space of L. We rely on this property and the finite dimensionality of $\mathcal{N}_{\mathrm{L}}$ to prove that $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ is a *bona fide* Banach space (see Theorem 5).

Having set the functional context, we now state our primary representer theorem.

THEOREM 1 (gTV optimality of splines for linear inverse problems). *Let us assume that the following conditions are met:*

1. *The regularization operator* $\mathrm{L} : \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \to \mathcal{M}(\mathbb{R}^d)$ *is spline-admissible in the sense of Definition 1.*
2. *The linear measurement operator* $\boldsymbol{\nu} : f \mapsto \boldsymbol{\nu}(f) = \big(\langle \nu_1, f\rangle, \ldots, \langle \nu_M, f\rangle\big)$ *maps* $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \to \mathbb{R}^M$ *and is weak*-continuous on* $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) = \big(C_{\mathrm{L}}(\mathbb{R}^d)\big)'$.
3. *The recovery problem is well-posed over the null space of* L*:* $\boldsymbol{\nu}(q_1) = \boldsymbol{\nu}(q_2) \Leftrightarrow q_1 = q_2$, *for any* $q_1, q_2 \in \mathcal{N}_{\mathrm{L}}$.

Then, the extremal points of the general constrained minimization problem

$$\beta = \min_{f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)} \|\mathrm{L}f\|_{\mathcal{M}} \quad s.t. \quad \boldsymbol{\nu}(f) \in \mathcal{C}, \tag{10}$$

where $\mathcal{C}$ *is any (feasible) convex compact subset of* $\mathbb{R}^M$, *are necessarily nonuniform* L*-splines of the form*

$$s(\boldsymbol{x}) = \sum_{k=1}^{K} a_k \rho_{\mathrm{L}}(\boldsymbol{x} - \boldsymbol{x}_k) + \sum_{n=1}^{N_0} b_n p_n(\boldsymbol{x}) \tag{11}$$

with parameters $K \leq M$, $\{\boldsymbol{x}_k\}_{k=1}^K$ *with* $\boldsymbol{x}_k \in \mathbb{R}^d$, $\mathbf{a} = (a_1, \ldots, a_K) \in \mathbb{R}^K$, *and* $\mathbf{b} = (b_1, \ldots, b_{N_0}) \in \mathbb{R}^{N_0}$. *Here,* $\{p_n\}_{n=1}^{N_0}$ *is a basis of* $\mathcal{N}_{\mathrm{L}}$ *and* $\mathrm{L}\{\rho_{\mathrm{L}}\} = \delta$ *so that* $\beta = \|\mathrm{L}s\|_{\mathcal{M}} = \sum_{k=1}^K |a_k| = \|\mathbf{a}\|_1$. *The full solution set of* (10) *is the convex hull of those extremal points.*

Theorem 1 is a powerful existence result that points towards the universality of nonuniform L-spline solutions. The key property here is $\mathrm{L}\{s\} = \sum_{k=1}^K a_k \delta(\cdot - \boldsymbol{x}_k)$, which follows from Conditions 1-3 in Definition 1 and is consistent with the more detailed characterization of splines presented in Section 5. For the time being, it suffices to remark that these splines are smooth (*i.e.*, infinitely differentiable) everywhere, except at their knot locations $\{\boldsymbol{x}_k\}$.

Although the extremal problem is defined over a continuum, the remarkable outcome is that the problem admits solutions that are intrinsically sparse, with the level of sparsity being measured by the minimum number K of required spline knots. In particular, this explains why the solution of a problem with a TV/L_1-type constraint on the derivative (resp., the second derivative) is piecewise-constant (resp., piecewise linear when $\mathrm{L} = \mathrm{D}^2$) with breakpoints at x_k. The other pleasing aspect is the direct connection between the functional concept of generalized TV and the ℓ_1-norm of the expansion coefficients $\mathbf{a}$.

We observe that the solution is made up of two components: an adaptive one that is specified by $\{\boldsymbol{x}_k\}$ and $\mathbf{a}$, and a linear regression term (with expansion coefficients $\mathbf{b}$) that describes the component in the null space of the operator. Since $\mathbf{b}$ does not contribute to $\|\mathrm{L}s\|_{\mathcal{M}}$, the optimization tends to maximize the contribution of the

null-space component. The main difficulty in finding the optimal solution is that K and $(\boldsymbol{x}_k)$ are problem-dependent and unknown *a priori*.

We have mentioned in Section 2.2 that the semi-norm $\|\mathrm{D}f\|_{\mathcal{M}}$ yields the classical total variation of a function in 1D. Unfortunately, there is no such direct connection for $d > 1$, the reason being that the multidimensional gradient $\boldsymbol{\nabla}$ is not spline-admissible because it is a vector operator. Instead, as a proxy for the popular total variation of Rudin and Osher [41], we suggest using the (fractional) Laplacian semi-norm $\|(-\Delta)^{\gamma/2} f\|_{\mathcal{M}}$ with $\gamma \geq d$, which is endowed with the same invariance and null-space properties. According to Theorem 1, such a γth-order regularization results in extremal points that are nonuniform polyharmonic splines [21, 37].

3.3. Connection with Unconstrained Problem. The statement in Theorem 1 is remarkably general. In particular, it covers the generic regularized least-squares problem

$$f_\lambda = \arg \min_{f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)} \left(\sum_{m=1}^{M} |y_m - \langle \nu_m, f \rangle|^2 + \lambda \|\mathrm{L}f\|_{\mathcal{M}} \right), \tag{12}$$

which is commonly used to formulate linear inverse/compressed-sensing problems [6, 9, 19, 23, 26]. The connection is obtained by taking $\mathcal{C} = \{\mathbf{z} \in \mathbb{R}^M : \|\mathbf{y} - \mathbf{z}\|_2^2 \leq \epsilon^2\} = B(\mathbf{y}; \epsilon)$, which is a ball of diameter ϵ centered on the measurement vector $\mathbf{y} = (y_1, \ldots, y_M)$. Indeed, since the data-fidelity term is (strictly) convex, the extreme points s_ϵ of (10) saturate the inequality such that $\|\mathbf{y} - \boldsymbol{\nu}(s_\epsilon)\|_2^2 = \epsilon^2$ and gTV is minimized with $\alpha = \alpha(\epsilon) = \|\mathrm{L}s_\epsilon\|_{\mathcal{M}}$. In the unconstrained form (12), the selection of a fixed $\lambda \in \mathbb{R}^+$ results in a particular value of the data error $\|\mathbf{y} - \boldsymbol{\nu}(f_\lambda)\|_2^2 = \epsilon'(\lambda)$ with the optimal solution $f_\lambda = s_{\epsilon'}$ having the same total variation as if we were looking at the primary problem (10) with $\mathcal{C} = B(\mathbf{y}; \epsilon')$.

To get further insights on the optimization problem (12), we can look at two limit cases. When $\lambda \to \infty$, the solution must take the form $f_\infty = p \in \mathcal{N}_{\mathrm{L}}$ so that $\|\mathrm{L}f_\infty\|_{\mathcal{M}} = 0$. It then follows that $\|\mathbf{y} - \boldsymbol{\nu}(f_\infty)\|_2^2 \leq \|\mathbf{y}\|^2 < \infty$. On the contrary, when $\lambda \to 0$, the minimization will force the data term $\|\mathbf{y} - \boldsymbol{\nu}(f_0)\|_2^2$ to vanish. Theorem 1 then ensures the existence of a nonuniform "interpolating" L-spline $f_0(\boldsymbol{x})$ with $\boldsymbol{\nu}(f_0) = \mathbf{y}$ and minimum gTV semi-norm.

3.4. Generalized Interpolation. In the latter interpolation scenario, the convex set $\mathcal{C}$ reduces to the single point $\mathcal{C} = \{\mathbf{y} \in \mathbb{R}^M\}$. This configuration is of special theoretical relevance because it enables us to refine our upper bound on the number K of spline knots.

THEOREM 2 (Generalized spline interpolant). *Under Assumptions 1-3 of Theorem 1, the extremal points of the (feasible) generalized interpolation problem*

$$\arg \min_{f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)} \|\mathrm{L}f\|_{\mathcal{M}} \quad \text{s.t.} \quad \boldsymbol{\nu}(f) = \mathbf{y} \tag{13}$$

are nonuniform L*-splines of the same form* (11) *as in Theorem 1, but with* $K \leq (M - N_0)$.

4. Application Areas. We first briefly comment on the admissibility conditions in Theorem 1 and indicate that the restrictions are minimal. To the best of our knowledge, the continuity of the measurement operator $\boldsymbol{\nu}$ is a necessary requirement for the mathematical analysis of any inverse problem. The difficulty here is that our native space $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) = (C_{\mathrm{L}}(\mathbb{R}^d))'$ is non-reflexive, which forces us to rely on the

weak*-topology. The continuity requirement in Theorem 1 is therefore equivalent to $\nu_m \in C_{\mathrm{L}}(\mathbb{R}^d)$ for $m = 1, \ldots, M$ where the Banach structure of the predual space $C_{\mathrm{L}}(\mathbb{R}^d)$ is laid out in Theorem 6. In particular, we refer to the norm inequality (25), which suggests that Condition 2 in Theorem 1 is met by picking $\nu_m \in L_{1,-n_0}(\mathbb{R}^d)$ where the latter is the Banach space associated with the weighted L_1-norm

$$\|f\|_{L_{1,-n_0}} = \int_{\mathbb{R}^d} |f(\boldsymbol{x})|(1 + \|\boldsymbol{x}\|)^{n_0} \mathrm{d}\boldsymbol{x}. \tag{14}$$

In fact, $L_{1,-n_0}(\mathbb{R}^d)$ is the predual of the space $L_{\infty,n_0}(\mathbb{R}^d)$ defined by (8), which implies that $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) = \big(C_{\mathrm{L}}(\mathbb{R}^d)\big)' \subset \big(L_{1,-n_0}(\mathbb{R}^d)\big)' = L_{\infty,n_0}(\mathbb{R}^d)$. The condition $\nu_m \in L_{1,-n_0}(\mathbb{R}^d)$ is a mild algebraic decay requirement that turns out to be satisfied by the impulse response of most physical devices. As for the requirement that the inverse problem is well defined over the null space of L (Condition 3), it a prerequisite to the success of any regularization scheme. Otherwise, there is simply no hope of turning an ill-posed problem into a well-posed one. For instance, in the introductory example with classical total-variation regularization, the constraint is that $\boldsymbol{\nu}$ should have at least one component ν_m such that $\langle \nu_m, 1 \rangle \neq 0$ which, again, is very mild requirement.

Next, we discuss examples of signal recovery that are covered by Theorems 1 and 2. The standard setting is that one is given a set of noisy measurements $\mathbf{y} = \boldsymbol{\nu}(s) +$ "noise" of an unknown signal s and that one is trying to recover s from $\mathbf{y}$ based on the solution of (12), or some variant of the problem involving some other (convex) data term—the most favorable choice being the log likelihood of the measurement noise. We shall then close the discussion section by briefly making the connection with a class of inverse problems in measure space; that is, the case L = Identity.

4.1. Interpolation. The task here is to reconstruct a continuous-domain signal from its (possibly noisy) nonuniform samples $\{s(\boldsymbol{x}_m)\}_{m=1}^M$, which is achieved by searching for the function $s : \mathbb{R}^d \to \mathbb{R}$ that fits the samples while minimizing $\|\mathrm{L}s\|_{\mathcal{M}}$. This corresponds to the problem setting in Theorem 1 with $\nu_m = \delta(\cdot - \boldsymbol{x}_m)$ and $\mathcal{C} = B(\mathbf{y}; \epsilon)$, where $\mathbf{y}$ denotes the measurement vector. Hence, the admissibility condition $\nu_m \in C_{\mathrm{L}}(\mathbb{R}^d)$ is equivalent to $(\mathrm{L}_{\boldsymbol{\phi}}^{-1})^*\{\delta(\cdot - \boldsymbol{x}_m)\} = g_{\boldsymbol{\phi}}(\boldsymbol{x}_m, \cdot) \in C_0(\mathbb{R}^d)$, where the boundedness is ensured by the stability condition in Theorem 4. The more technical continuity requirement is achieved when ρ_{L} is continuous (Hölder exponent $r_0 > 0$). This happens when the order of the differential operator is greater than one, which seems to exclude[2] simple operators such as D (piecewise-constant approximation). This limitation notwithstanding, our theoretical results are directly applicable to the problems of adaptive regression splines [38] with $\mathrm{L} = \mathrm{D}^N$, the construction of shape-preserving splines [35], as well as a whole range of variations including TV denoising.

4.2. Generalized Sampling. The setting is analogous to the previous one, except that the samples are now observed through a sampling aperture $\phi \in L_{1,-n_0}(\mathbb{R}^d)$ so that $\nu_m = \phi(\cdot - \boldsymbol{x}_m)$ [24, 50]. The function ϕ may, for example, correspond to the point-spread function of a microscope. Then, the recovery problem is equivalent to a deconvolution [18]. Since the measurements are obtained by integration of s against an ordinary function $\nu_m \in L_{1,-n_0}(\mathbb{R}^d)$, there is no requirement for the continuity of ρ_{L} because of the implicit smoothing effect of ϕ. This means that essentially no restrictions apply.

[2]We can bypass this somewhat artificial limitation by replacing the ideal sampler by a quasi-ideal sampling device that involves a mollified version of a Dirac impulse.

4.3. Compressed Sensing. The result of Theorem 1 is highly relevant to compressed sensing, especially since the underlying L_1/TV signal-recovery problem is formulated in the continuous domain. We like to view (11) as the prototypical form of a piecewise-smooth signal that is intrinsically sparse with sparsity $K = \|\boldsymbol{a}\|_0$. The model also conforms with the notion of a finite rate of innovation [56]. If we know that the unknown signal s has such a form, then Theorem 1 suggests that we can attempt to recover it from an M-dimensional linear measurement $\mathbf{y} = \boldsymbol{\nu}(s)$ by solving the optimization problem (10) with $\mathcal{C} = B(\boldsymbol{y}; \epsilon)$, which is in agreement with the predominant paradigm in the field. While the theorem states that $M \geq K$, common sense dictates that we should take $M > N_{\text{freedom}}$, where $N_{\text{freedom}} = 2K + N_0$ is the number of degrees of freedom of the underlying model. The difficulty, of course, is that a subset of those parameters (the spline knots $\boldsymbol{x}_k$) induce a model dependency that is highly nonlinear.

4.4. Inverse Problems in the Space of Measures. Some of the theoretical results of this paper are also of direct relevance for inverse problems that are formulated in the space $\mathcal{M}(\mathbb{R}^d)$ of measures [5]. The prototypical example is the recovery of the location (with super-resolution precision) of a series of Dirac impulses from noisy measurements, which may be achieved through the continuous-domain minimization of the total variation of the underlying measure [8, 16, 22, 25]. The Fisher-Jerome theorem [27, Theorem 1] as well as our extension for the unbounded domain $\mathbb{R}^d$ and arbitrary convex sets (Theorem 7) support this kind of algorithm, as they guarantee the existence of sparse solutions—understood as a sum of Dirac spikes—for this family of problems.

5. Splines and Operators. We now switch to the explanatory part of the presentation. The first important concept that is implicit in the statement of Theorems 1 and 2 is the powerful association between splines and operators, the idea being that the selection of an admissible operator L specifies a corresponding type of splines [46, 47][55, Chapter 6].

Sorted by increasing complexity, the three types of operators that are of relevance to us are: (i) ordinary differential operators, which are polynomials of the derivative operator $\mathrm{D} = \frac{\mathrm{d}}{\mathrm{d}x}$ [13, 46, 52]; (ii) partial differential operators such as the Laplacian Δ (or some polynomial thereof); and (iii) fractional derivatives such as D^γ or $(-\Delta)^{\frac{\gamma}{2}}$ with $\gamma \in \mathbb{R}^+$ whose Fourier symbols are $(\mathrm{j}\omega)^\gamma$ and $\|\boldsymbol{\omega}\|^\gamma$, respectively [21, 51, 53]. It can be shown that all linear-shift-invariant operators of Type (i) and all elliptic operators of Type (ii) are spline-admissible in the sense of Definition 1. This is also known to be true for fractional derivatives and fractional Laplacians with $\gamma \geq d$ [21, 51].

Let us mention that the issue of making sure that the null space of the operator L is finite-dimensional is often nontrivial for $d > 1$. It is a fundamental aspect that is addressed in the L_2 theory of radial basis functions and polyharmonic splines with the definition of the appropriate native spaces [58, Chapter 10]. Here, we have chosen to bypass some of these technicalities by including a growth restriction (*i.e.*, the condition that $q \in L_{\infty,n_0}(\mathbb{R}^d)$) in the definition of $\mathcal{N}_{\mathrm{L}}$. A fundamental property in that respect is that the finite-dimensional null space of a LSI operator can only include exponential polynomial components of the form $\boldsymbol{x}^{\boldsymbol{m}} \mathrm{e}^{\mathrm{j}\langle \boldsymbol{\omega}_0, \boldsymbol{x} \rangle}$, which correspond to a zero of multiplicity at least $|\boldsymbol{m}| + 1$ of the frequency response $\widehat{L}(\boldsymbol{\omega})$ at $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ (see [55, Proposition 6.1 p. 118] and [32, Section 6]).

Once it is established that L is spline-admissible, one can rely on the following unifying distributional definition of a spline.

DEFINITION 2 (Nonuniform L-spline). *A function $s : \mathbb{R}^d \to \mathbb{R}$ of slow growth (i.e., $s \in L_{\infty,n_0}(\mathbb{R}^d)$ with $n_0 \geq 0$) is said to be a nonuniform* L*-spline if*

$$\mathrm{L}\{s\} = \sum_k a_k \delta(\cdot - \boldsymbol{x}_k) = w_\delta, \tag{15}$$

where (a_k) is a sequence of weights and the Dirac impulses are located at the spline knots $\{\boldsymbol{x}_k\}$.

The generalized function $\mathrm{L}\{s\} = w_\delta$ is called the innovation of the spline because it contains the crucial information for its description: the positions $\{\boldsymbol{x}_k\}$ of the knots and the amplitudes (a_k) of the corresponding discontinuities.

The one-dimensional brands of greatest practical interest are the polynomial splines with $\mathrm{L} = \mathrm{D}^m$ [15, 45] and the exponential splines [13, 46, 52] with $\mathrm{L} = c_m\mathrm{D}^m + \cdots + c_1\mathrm{D} + c_0\mathrm{I}$, where $\mathrm{I} = \mathrm{D}^0$ denotes the identity operator. Their multidimensional counterparts are the polyharmonic splines with $\mathrm{L} = (-\Delta)^{\gamma/2}$ [21, 37] and the Sobolev splines with $\mathrm{L} = (\mathrm{I} - \Delta)^{\gamma/2}$ for $\gamma \geq d$ [57]. The connection with the theory of Sobolev spaces is that the Green's functions of $(-\Delta)^{\gamma/2}$ (resp., $(\mathrm{I} - \Delta)^{\gamma/2}$) are the kernels of the Riesz (resp., Bessel) potentials [31].

For a constructive use of Definition 2, we also need to be able to re-synthesize the spline s from its innovation. In the case of our introductory example with $\mathrm{L} = \mathrm{D}$ (see Figure 1), one simply integrates w_δ, which yields (6) (up to the integration constant b_1) owing to the property that $\mathrm{D}^{-1}\{\delta(\cdot - x_k)\}(x) = \int_{-\infty}^{x} \delta(\tau - x_k)\mathrm{d}\tau = \mathbb{1}_+(x - x_k)$. In principle, the same inversion procedure is applicable for the generic operator L and amounts to substituting the δ distribution in (15) by the Green's function ρ_L. The only delicate part is the proper handling of the "integration constants" (the part of the solution that lies in the null space of the operator), which is achieved through the specification of N_0 linear boundary conditions of the form $\langle \phi_n, s \rangle = 0$.

We now show that the underlying functionals $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_{N_0})$ can be incorporated in the specification of an appropriate right-inverse operator $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$. Our construction requires that $\boldsymbol{\phi}$ first be matched to a basis of $\mathcal{N}_\mathrm{L}$ such as to form a biorthogonal system. We note that this is always feasible as long as the ϕ_n are linearly independent with respect to $\mathcal{N}_\mathrm{L}$. (An explicit construction is given in the proof of Theorem 2.)

DEFINITION 3. *The pair $(\boldsymbol{\phi}, \boldsymbol{p}) = (\phi_n,\ p_n)_{n=1}^{N_0}$ is called a biorthogonal system for $\mathcal{N}_\mathrm{L} \subset \mathcal{M}_\mathrm{L}(\mathbb{R}^d)$ if $\{p_n\}_{n=1}^{N_0}$ is a basis of $\mathcal{N}_\mathrm{L}$ and the vector of "boundary" functionals $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_{N_0})$ with $\phi_n \in \mathcal{N}_\mathrm{L}'$ satisfy the biorthogonality condition $\boldsymbol{\phi}(p_n) = \mathbf{e}_n$ where $\mathbf{e}_n$ is the nth element of the canonical basis.*

The interest of such a system is that any $q \in \mathcal{N}_\mathrm{L}$ has a unique representation as $q = \sum_{n=1}^{N_0} \langle \phi_n, q \rangle p_n$ with associated norm $\|\boldsymbol{\phi}(q)\|_2$.

The fundamental requirement for our formulation is the stability/continuity of the inverse operator $\mathrm{L}_{\boldsymbol{\phi}}^{-1} : \mathcal{M}(\mathbb{R}^d) \to \mathcal{M}_\mathrm{L}(\mathbb{R}^d)$. Since $\mathcal{M}_\mathrm{L}(\mathbb{R}^d) \subset L_{\infty,n_0}(\mathbb{R}^d)$ by construction, we can control stability by relying on Theorem 3 whose proof is given in Appendix A.

THEOREM 3. *The generic linear operator $\mathrm{G} : w \mapsto f = \int_{\mathbb{R}^d} g(\cdot, \boldsymbol{y}) w(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}$ continuously maps $\mathcal{M}(\mathbb{R}^d) \to L_{\infty,\alpha_0}(\mathbb{R}^d)$ with $\alpha_0 \in \mathbb{R}$ if and only if its kernel g is measurable and*

$$\operatorname*{ess\,sup}_{\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^d} \left(|g(\boldsymbol{x}, \boldsymbol{y})|\ (1 + \|\boldsymbol{x}\|)^{-\alpha_0} \right) < \infty. \tag{16}$$

This allows us to characterize the desired operator in term of its Schwartz' kernel (or generalized impulse response) $g_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{\delta(\cdot - \boldsymbol{y})\}(\boldsymbol{x})$.

THEOREM 4 (Stable right-inverse of L). *Let $(\phi_n,\ p_n)_{n=1}^{N_0}$ be a biorthogonal system for $\mathcal{N}_{\mathrm{L}} \subset \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \subset L_{\infty,n_0}(\mathbb{R}^d)$. Then, there exists a unique operator $\mathrm{L}_{\boldsymbol{\phi}}^{-1} : \varphi \mapsto \mathrm{L}_{\boldsymbol{\phi}}^{-1}\varphi = \int_{\mathbb{R}^d} g_{\boldsymbol{\phi}}(\cdot,\boldsymbol{y})\varphi(\boldsymbol{y})\mathrm{d}\boldsymbol{y}$ such that*

$$\mathrm{L}\mathrm{L}_{\boldsymbol{\phi}}^{-1}\varphi = \varphi \qquad (\textit{right-inverse property}) \tag{17}$$

$$\boldsymbol{\phi}(\mathrm{L}_{\boldsymbol{\phi}}^{-1}\varphi) = \mathbf{0} \qquad (\textit{boundary condidions}) \tag{18}$$

for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$. The kernel of this operator is

$$g_{\boldsymbol{\phi}}(\boldsymbol{x},\boldsymbol{y}) = \rho_{\mathrm{L}}(\boldsymbol{x}-\boldsymbol{y}) - \sum_{n=1}^{N_0} p_n(\boldsymbol{x})q_n(\boldsymbol{y}), \tag{19}$$

with ρ_{L} such that $\mathrm{L}\rho_{\mathrm{L}} = \delta$ and $q_n(\boldsymbol{y}) = \langle \phi_n, \rho_{\mathrm{L}}(\cdot - \boldsymbol{y})\rangle$. Moreover, if $g_{\boldsymbol{\phi}}$ satisfies the stability condition (16) with $\alpha_0 = n_0$, then $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ admits a continuous extension $\mathcal{M}(\mathbb{R}^d) \to L_{\infty,n_0}(\mathbb{R}^d)$ with (17) and (18) remaining valid for all $\varphi \in \mathcal{M}(\mathbb{R}^d)$.

The proof of Theorem 4 is given in Appendix B.

Since the choice of the N_0 linear boundary functionals ϕ_n is essentially arbitrary, there is flexibility in defining admissible inverse operators. The important ingredient for our formulation is the existence of such inverses with the unconditional guarantee of their stability (see Theorem 5 below).

To put this result into context, we now provide some illustrative examples. For $\mathrm{L} = \mathrm{D}^{N_0}$, we have that $n_0 = (N_0 - 1)$, $\rho_{\mathrm{D}^{N_0}}(x) = \frac{x_+^{n_0}}{n_0!}$, and $p_n(x) = \frac{x^{n-1}}{(n-1)!}$ for $n = 1,\dots,N_0$, where the polynomial basis is biorthogononal to $\boldsymbol{\phi}$ with $\phi_n(x) = (-1)^{(n-1)}\delta^{(n-1)}(x)$. This (canonical) choice of boundary functionals then translates into the construction of an inverse operator $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ that imposes the vanishing of the function and its derivatives at the origin. By applying (19) and recognizing the binomial expansion of $(x-y)^{n_0}$, we simplify the expression of the kernel of this operator as

$$\begin{aligned} g_{\boldsymbol{\phi}}(x,y) &= \frac{(x-y)_+^{n_0}}{n_0!} - \sum_{n=0}^{n_0} \frac{x^n}{n!}\frac{(-y)_+^{n_0-n}}{(n_0-n)!} \\ &= \begin{cases} \frac{(x-y)^{n_0}}{n_0!}\mathbb{1}_{(0,x]}(y), & x \geq 0 \\ -\frac{(x-y)^{n_0}}{n_0!}\mathbb{1}_{(x,0]}(y), & x < 0. \end{cases} \end{aligned}$$

The crucial observation here is that the function $g_{\boldsymbol{\phi}}(x,\cdot)$ with $x \in \mathbb{R}$ fixed is compactly supported and bounded. Moreover, $\|g_{\boldsymbol{\phi}}(x,\cdot)\|_\infty = |g_{\boldsymbol{\phi}}(x,0)| = \frac{x^{n_0}}{n_0!}$ so that $g_{\boldsymbol{\phi}}$ obviously satisfies the stability bound (16) with $\alpha_0 = n_0$. By contrast, the condition fails for the conventional shift-invariant inverse $\varphi \mapsto \rho_{\mathrm{D}^{N_0}} * \varphi$ (n_0-fold integrator), which stresses out the non-trivial stabilizing effect of the second correction term in (19). The other important consequence of the correction is the vanishing of $g_{\boldsymbol{\phi}}(x,y)$ as $y \to \pm\infty$ for any fixed $x \in \mathbb{R}^d$, contrary to its leading term $(x-y)_+^{n_0}/n_0!$ which does not decay (and even grows) as $y \to -\infty$.

The primary usage of the inverse operators of Theorem 4 is the resolution of differential equations of the form

$$\mathrm{L}s = w \quad \text{s.t.} \quad \boldsymbol{\phi}(s) = (b_1,\dots,b_{N_0}) \tag{20}$$

for some $w \in \mathcal{M}(\mathbb{R}^d)$. Indeed, by invoking the properties of $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ and the biorthogonality of $(\boldsymbol{\phi},\boldsymbol{p})$, we readily show that (20) admits a unique solution in $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$, which

is given by

$$s = \mathrm{L}_{\boldsymbol{\phi}}^{-1} w + \sum_{n=1}^{N_0} b_n p_n.$$

For the particular case of the spline innovation w_δ in Definition 2, we find that

$$s = \sum_k a_k \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{\delta(\cdot - \boldsymbol{x}_k)\} + \sum_{n=1}^{N_0} b_n p_n$$

which, upon substitution of the kernel given by (19), results in a form that is the same as (11) in Theorem 1 modulo some adjustment of the constants b_n.

6. Native or Generalized Beppo-Levi Spaces. The search for the solution of our optimization problem is performed over the native space $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ defined by (9), which is the largest space over which our gTG regularization functional is well defined. The delicate aspect is that $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ is specified in terms of a semi-norm, in analogy with the definition of the classical Beppo-Levi spaces of order $n \in \mathbb{N}$ and exponent $p \geq 1$, written as $\mathcal{B}_{p,n}(\mathbb{R}^d) = \{f \in \mathcal{S}'(\mathbb{R}^d) : \|\partial^{\boldsymbol{m}} f\|_{L_p} < \infty \text{ for all multi-indices } |\boldsymbol{m}| = n\}$ [17, 34]. Hence, in 1D, the proposed definition of $\mathcal{M}_{\mathrm{D}^n}(\mathbb{R})$ is a slight extension of $\mathcal{B}_{1,n}(\mathbb{R})$. In higher dimensions, it can be shown[3] that $\mathcal{B}_{p,2n}(\mathbb{R}^d) = \{f \in L_{\infty,2n-1}(\mathbb{R}^d) : \|(-\Delta)^n f\|_{L_p} < \infty\}$, so that there also exists a close connection between $\mathcal{B}_{1,2n}(\mathbb{R}^d)$ and $\mathcal{M}_{(-\Delta)^n}(\mathbb{R}^d)$.

The crucial point for our formulation is that $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ also happens to be a complete normed (or Banach) space when equipped with the proper direct-sum topology. We shall now make this structure explicit with the help of the inverse operators defined in Theorem 4. Since the principle is similar to the characterization of the Beppo-Levi spaces, we shall also refer to $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ as a generalized Beppo-Levi space.

THEOREM 5 (Banach-space structure of native space). *Let* L *be a spline-admissible operator, $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ its native space defined by (9), and $(\boldsymbol{\phi}, \boldsymbol{p})$ some biorthogonal system for its null space $\mathcal{N}_{\mathrm{L}}$. Then, the following equivalent conditions hold:*

1. *The right-inverse operator $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ specified by Theorem 4 isometrically maps $\mathcal{M}(\mathbb{R}^d) \to \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \subset L_{\infty,n_0}(\mathbb{R}^d)$, while its kernel necessarily fullfills the stability condition*

$$C_{\boldsymbol{\phi}} = \sup_{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d} \left(|g_{\boldsymbol{\phi}}(\boldsymbol{x},\boldsymbol{y})|\,(1+\|\boldsymbol{x}\|)^{-n_0}\right) < \infty. \tag{21}$$

2. *Every $f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ admits a unique representation as*

$$f = \mathrm{L}_{\boldsymbol{\phi}}^{-1} w + p,$$

where $w = \mathrm{L}\{f\} \in \mathcal{M}(\mathbb{R}^d)$ and $p = \sum_{n=1}^{N_0} \langle f, \phi_n\rangle p_n \in \mathcal{N}_{\mathrm{L}}$.

3. *$\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ is a Banach space equipped with the norm*

$$\|f\|_{\mathrm{L},\boldsymbol{\phi}} = \|\mathrm{L}f\|_{\mathcal{M}} + \|\boldsymbol{\phi}(f)\|_2. \tag{22}$$

Proof. As preparation, we define a subset of $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ as

$$\mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d) = \{f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) : \boldsymbol{\phi}(f) = \mathbf{0}\}. \tag{23}$$

[3]The argument is that the only functions that are harmonic and of slow growth are polynomials.

Since the boundary conditions $\boldsymbol{\phi}(f) = \mathbf{0}$ are linear, $\mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d)$ is clearly a vector space. We now show that it is a Banach space when equipped with the norm $\|\cdot\|_{\mathrm{L}} = \|\mathrm{L}\{\cdot\}\|_{\mathcal{M}}$. By definition, $\|\cdot\|_{\mathrm{L}}$ is a semi-norm on $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$, meaning that it fulfills the properties of a norm, except for the unicity condition. To establish the latter on $\mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d)$, we consider $f \in \mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d)$ such that $\|f\|_{\mathrm{L}} = 0$, which is equivalent to $f \in \mathcal{N}_{\mathrm{L}}$. Since $f \in \mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d)$ (by hypothesis), we have that $\boldsymbol{\phi}(f) = \mathbf{0}$, from which we deduce that $f = \sum_{n=1}^{N_0} \langle \phi_n, f \rangle p_n = 0$, as expected. This proves that $\mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d)$ is isometrically isomorphic to $\mathcal{M}(\mathbb{R}^d)$ and, hence, a Banach space. Alternatively, one can also view $\mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d)$ as a concrete transcription (or representative within the equivalence class) of the abstract quotient space $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)/\mathcal{N}_{\mathrm{L}}$.

1. Existence and Stability of Inverse Operators. We have just revealed that L is a bijective, norm-preserving mapping $\mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d) \to \mathcal{M}(\mathbb{R}^d)$. This allows us to invoke the bounded-inverse theorem, which ensures the existence and boundedness (here, an isometry) of the inverse operator $\mathrm{L}^{-1} : \mathcal{M}(\mathbb{R}^d) \to \mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d)$. The relevant L^{-1} is precisely the unique operator $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ identified in Theorem 4, as it imposes the boundary condition $\boldsymbol{\phi}(\mathrm{L}_{\boldsymbol{\phi}}^{-1} w) = \mathbf{0}$ for all $w \in \mathcal{M}(\mathbb{R}^d)$. Finally, we use the same technique as in the proof of Theorem 3 to establish the necessity of the stability condition (16) with $\alpha_0 = n_0$.

2. Direct Sum Decomposition. Since the system $(\boldsymbol{\phi}, \boldsymbol{p})$ is biorthogonal, the operator $\mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}} : f \mapsto \sum_{n=1}^{N_0} \langle \phi_n, f \rangle p_n$ is a continuous projection operator $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \to \mathcal{N}_{\mathrm{L}}(\mathbb{R}^d)$. It follows that any element $f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ has a unique decomposition as $f = f_1 + q$, where $q = \mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}}\{f\} \in \mathcal{N}_{\mathrm{L}}$ and $f_1 = (f - q)$ with $\boldsymbol{\phi}(f_1) = \mathbf{0}$. This last condition implies that $f_1 \in \mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d)$ so that f_1 has a unique representation as $f_1 = \mathrm{L}_{\boldsymbol{\phi}}^{-1} w$, where $w = \mathrm{L} f_1 = \mathrm{L} f \in \mathcal{M}(\mathbb{R}^d)$. Since $\mathcal{M}_{\mathrm{L},\boldsymbol{\phi}} \cap \mathcal{N}_{\mathrm{L}} = \{0\}$, this expresses the structural property that $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) = \mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d) \oplus \mathcal{N}_{\mathrm{L}}$.

3. Identification of the Underlying Norm. Any element $p \in \mathcal{N}_{\mathrm{L}}$ is uniquely characterized by its expansion coefficients $\boldsymbol{\phi}(p)$ in the basis $\boldsymbol{p}$. The same holds true for $q = \mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}}\{f\} \in \mathcal{N}_{\mathrm{L}}$ with $\boldsymbol{\phi}(q) = \boldsymbol{\phi}(f)$ for any $f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$. Since $\mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d)$ and $\mathcal{N}_{\mathrm{L}}$ are both Banach spaces, we can equip their direct sum $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ with the composite norm $\|f\|_{\mathrm{L},\boldsymbol{\phi}} = \|w\|_{\mathcal{M}} + \|\boldsymbol{\phi}(f)\|_2$, with the guarantee that the Banach-space property is preserved.

For the converse implication, we simply identify $\mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d)$ as the closed subspace of $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ with the property that $\|f\|_{\mathrm{L},\boldsymbol{\phi}} = \|\mathrm{L} f\|_{\mathcal{M}}$. ∎

The connection with the L-spline s of Definition 2 is that $s \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ if and only if the ℓ_1-norm of its spline weights $\mathbf{a} = (a_1, \ldots, a_K)$ is finite. Indeed, we have that $\|\mathrm{L} s\|_{\mathcal{M}} = \|w_\delta\|_{\mathcal{M}} = \sum_{k=1}^{K} |a_k| = \|\mathbf{a}\|_{\ell_1}$, owing to the property that $\|\delta(\cdot - \boldsymbol{x}_k)\|_{\mathcal{M}} = 1$.

We note that the choice of gTV is essential here since the simpler (and *a priori* only slightly more restrictive) L_1-norm regularization $\|\mathrm{L} s\|_{L_1}$ would exclude the spline solutions that are of interest to us because $\delta \notin L_1(\mathbb{R}^d)$.

Our final ingredient is the identification of the predual space of $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$, which is denoted by $C_{\mathrm{L}}(\mathbb{R}^d)$.

THEOREM 6 (Predual of native space). *Let $(\boldsymbol{\phi}, \boldsymbol{p})$ be a biorthogonal system of $\mathcal{N}_{\mathrm{L}} \subset L_{\infty,n_0}(\mathbb{R}^d)$ and $C_{\mathrm{L},\boldsymbol{p}}(\mathbb{R}^d)$ be the image of $C_0(\mathbb{R}^d)$ by $\mathrm{L}^* : C_0(\mathbb{R}^d) \to C_{\mathrm{L},\boldsymbol{p}}(\mathbb{R}^d)$. Then, $\mathcal{M}_{\mathrm{L}} = \big(C_{\mathrm{L}}(\mathbb{R}^d)\big)'$ where $C_{\mathrm{L}}(\mathbb{R}^d) = C_{\mathrm{L},\boldsymbol{p}}(\mathbb{R}^d) \oplus \mathcal{N}'_{\mathrm{L}}$ with $\mathcal{N}'_{\mathrm{L}} = \mathrm{span}\{\phi_n\}_{n=1}^{N_0}$. $C_{\mathrm{L}}(\mathbb{R}^d)$ is a Banach space equipped with the norm*

$$\|f\|'_{\mathrm{L},\boldsymbol{\phi}} = \|\mathrm{L}_{\boldsymbol{\phi}}^{-1*} f\|_\infty + \|\boldsymbol{p}(f)\|_2. \tag{24}$$

where $\mathrm{L}_{\boldsymbol{\phi}}^{-1*} = \left(\mathrm{L}_{\boldsymbol{\phi}}^{-1}\right)^*$ *is the adjoint of* $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$. *Moreover, there exists a constant* $C > 0$ *such that*

$$\|f\|'_{\mathrm{L},\boldsymbol{\phi}} \leq C\|f\|_{L_{1,-n_0}} \tag{25}$$

for any $f \in L_{1,-n_0}(\mathbb{R}^d)$.

The proof is given in Appendix C. The direct-sum decomposition in Theorem 6 is achieved by means of the operator $\mathrm{Proj}_{\mathcal{N}'_{\mathrm{L}}} : f \mapsto q = \sum_{n=1}^{N_0} \langle p_n, f \rangle \phi_n$ with $\|q\| = \|\boldsymbol{p}(f)\|_2 = \|\boldsymbol{p}(q)\|_2$, which relies on the biorthogonality of $(\boldsymbol{\phi}, \boldsymbol{p})$ to project $\mathcal{C}_{\mathrm{L}}(\mathbb{R}^d)$ onto $\mathcal{N}'_{\mathrm{L}}(\mathbb{R}^d)$. This also means that $C_{\mathrm{L},\boldsymbol{p}}(\mathbb{R}^d)$ can be defined as $C_{\mathrm{L},\boldsymbol{p}}(\mathbb{R}^d) = \{f \in C_{\mathrm{L}}(\mathbb{R}^d) : \boldsymbol{p}(f) = \mathbf{0}\}$, in direct analogy with the definition of $\mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d)$ in (23).

7. Proof of Theorems 1 and 2. Our technique of proof will be to first establish the optimality of innovation-type solutions of the form that appear in Definition 2 for general linear inverse problems defined on $\mathcal{M}(\mathbb{R}^d)$ (Theorem 7) and to then transfer the result to $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ with the help of the stable inverse operators specified in Theorem 4. The first step is achieved by generalizing an earlier result by Fisher and Jerome [27].

Let $\mathcal{H}$ be the direct sum of $\mathcal{M}(\mathbb{R}^d) = \left(C_0(\mathbb{R}^d)\right)'$ and a finite-dimensional space $\mathcal{N}$ equipped with some norm $\|\cdot\|_{\mathcal{N}}$. The generic element of $\mathcal{H}$ is $f = (w, p)$ with $\|f\|_{\mathcal{H}} = \|w\|_{\mathcal{M}} + \|p\|_{\mathcal{N}}$.

THEOREM 7 (Generalized Fisher-Jerome theorem). *Let* $\boldsymbol{F} : \mathcal{H} \to \mathbb{R}^M$ *with* $M \geq N_0 = \dim(\mathcal{N})$ *be a weak*-continuous linear map such that*

$$B\|p\|_{\mathcal{N}} \leq \|\boldsymbol{F}(0, p)\|_2 \tag{26}$$

for some constant $B > 0$ *and every* $p \in \mathcal{N}$. *Let* $\mathcal{C}$ *be a convex compact subset of* $\mathbb{R}^M$ *such that* $\mathcal{U} = \boldsymbol{F}^{-1}(\mathcal{C}) = \{(w, p) \in \mathcal{H} : \boldsymbol{F}(w, p) \in \mathcal{C}\}$ *is nonempty (feasibility hypothesis). Then,*

$$\mathcal{V} = \arg \min_{(w,p)\in\mathcal{U}} \|w\|_{\mathcal{M}}$$

is a nonempty, convex, weak-compact subset of* $\mathcal{H}$ *with extremal points* (w_δ, p) *of the form*

$$w_\delta = \sum_{k=1}^{K} a_k \delta(\cdot - \boldsymbol{x}_k) \tag{27}$$

with $K \leq M$ *and* $\boldsymbol{x}_k \in \mathbb{R}^d$ *for* $k = 1, \ldots, K$, *and* $\min_{(w,p)\in\mathcal{U}} \|w\|_{\mathcal{M}} = \sum_{k=1}^{K} |a_k|$.

Theorem 7 is the most technical component of our formulation as it involves the weak* topology. The details of the proof are laid out in Appendix D together with a precise definition of the underlying concepts.

The essence of Theorem 7 is very similar to Fisher-Jerome's original result [27, Theorem 1], except for two crucial points: (i) the fact that they are only considering measures defined over a bounded domain $\Omega \subset \mathbb{R}^d$ (or, by extension, on a compact metric space), and (ii) the nature of the constraints which, in their case, is limited to coordinatewise inequalities of the form $z_{1,m} \leq [\boldsymbol{F}(w, p)]_m \leq z_{2,m}$. These differences are substantial enough to justify a new, self-contained proof. In particular, we believe that our extension for functions defined on $\mathbb{R}^d$ (beyond the compact Hausdorff framework of [27]) is essential for covering nonlocal operators such as fractional derivatives, and for deploying Fourier-domain/signal-processing techniques.

Our primary constraint for the validity of Theorem 7 is the existence of the lower bound (26). We now show that this property is implicit in the statement of the hypotheses of Theorem 1.

PROPOSITION 8. *Let $(\phi_n, p_n)_{n=1}^{N_0}$ be a biorthogonal system of $\mathcal{N}_{\mathrm{L}} \subset \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ such that $q = \sum_{n=1}^{N_0} \langle \phi_n, q \rangle p_n$ for all $q \in \mathcal{N}_{\mathrm{L}}$. Then, Condition 3 in Theorem 1 is equivalent to the existence of a constant $0 < B$ such that*

$$B\|q\|_{\mathcal{N}_{\mathrm{L}}} \leq \|\boldsymbol{\nu}(q)\|_2, \quad \forall q \in \mathcal{N}_{\mathrm{L}} \tag{28}$$

with $\|q\|_{\mathcal{N}_{\mathrm{L}}}^2 = \|\boldsymbol{\phi}(q)\|_2^2 = \sum_{n=1}^{N_0} |\langle \phi_n, q \rangle|^2$.

While there are softer ways of establishing this equivalence, we have chosen an explicit approach that also serves as background for the proof of Theorem 2.

Proof. Any $q \in \mathcal{N}_{\mathrm{L}}$ has a unique expansion $q = \sum_{n=1}^{N_0} c_n p_n$ with $\mathbf{c} = \boldsymbol{\phi}(q)$ and $\|q\|_{\mathcal{N}_{\mathrm{L}}} = \|\mathbf{c}\|_2$. The property that q is uniquely determined by its measurements $\mathbf{b} = \boldsymbol{\nu}(q)$ is therefore equivalent to $\mathbf{c}$ also being the solution of the overdetermined system $\mathbf{P}\mathbf{c} = \mathbf{b}$ with

$$\mathbf{P} = [\boldsymbol{\nu}(p_1) \ \cdots \ \boldsymbol{\nu}(p_{N_0})]. \tag{29}$$

It is well known that such a system is solvable if and only if $(\mathbf{P}^T\mathbf{P})$ is invertible and that its (least-squares) solution is given by

$$\mathbf{c} = (\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{b}.$$

This characterization then yields the norm estimate

$$\|q\|_{\mathcal{N}_{\mathrm{L}}} = \|\mathbf{c}\|_2 \leq \frac{\sigma_{\max}(\mathbf{P})}{\sigma_{\min}^2(\mathbf{P})} \|\boldsymbol{\nu}(q)\|_2,$$

where $\sigma_{\min}(\mathbf{P}) = \sigma_{\min}(\mathbf{P}^T)$ and $\sigma_{\max}(\mathbf{P})$ are the minimum and maximum singular values of $\mathbf{P}$, respectively. Finally, the invertibility of $(\mathbf{P}^T\mathbf{P})$ is equivalent to $\sigma_{\min}^2(\mathbf{P}) = \lambda_{\min}(\mathbf{P}^T\mathbf{P}) > 0$, while the continuity assumption on $\boldsymbol{\nu}$ ensures that $\sigma_{\max}(\mathbf{P}) < \infty$. The constant is then given by $B = \sigma_{\min}^2(\mathbf{P})/\sigma_{\max}(\mathbf{P})$. ∎

Proof of Theorem 1. Let $(\boldsymbol{\phi}, \boldsymbol{p})$ be a biorthogonal system for $\mathcal{N}_{\mathrm{L}}$. Then, by Theorem 5, any function $f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ has a unique decomposition as $f = \mathrm{L}_{\boldsymbol{\phi}}^{-1} w + p$ with $w = \mathrm{L}f \in \mathcal{M}(\mathbb{R}^d)$ and $p \in \mathcal{N}_{\mathrm{L}}$. This allows us to interpret the measurement process $f \mapsto \boldsymbol{\nu}(f) = \langle \boldsymbol{\nu}, f \rangle$ as the linear map $\boldsymbol{F} : \mathcal{H} \to \mathbb{R}^M$ such that

$$\begin{aligned} \langle \boldsymbol{\nu}, f \rangle &= \langle \boldsymbol{\nu}, \mathrm{L}_{\boldsymbol{\phi}}^{-1} w \rangle + \langle \boldsymbol{\nu}, p \rangle \\ &= \langle \mathrm{L}_{\boldsymbol{\phi}}^{-1*} \boldsymbol{\nu}, w \rangle + \langle \boldsymbol{\nu}, p \rangle = \boldsymbol{F}(w, p). \end{aligned}$$

We also know from Theorem 6 that L_{ϕ}^{-1*} is an isometry $C_{\mathrm{L}}(\mathbb{R}^d) \to C_0(\mathbb{R}^d)$. Hence, the weak*-continuity of $\boldsymbol{\nu} : \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \to \mathbb{R}^M$ is equivalent to the weak*-continuity of $\boldsymbol{F} : \mathcal{H} \to \mathbb{R}^M$. The complementary lower bound is given by Proposition 8 as

$$B\|p\|_{\mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}} \leq \|\boldsymbol{\nu}(p)\|_2 = \|\boldsymbol{F}(0, p)\|_2.$$

With this new representation, the constrained minimization problem is equivalent to the one considered in Theorem 7 with $\mathcal{N} = \mathcal{N}_{\mathrm{L}}$, which ensures that all extreme points of the solution set are of the form (p, w_δ) with $w_\delta = \sum_{k=1}^{K} a_k \delta(\cdot - \boldsymbol{x}_k)$, $K \leq M$,

and $\boldsymbol{x}_k \in \mathbb{R}^d$. Upon application of the (stable) right-inverse operator, this maps into $s = \mathrm{L}_{\boldsymbol{\phi}}^{-1} w_\delta + p$, where p is a suitable component that is in the null space of the operator. Finally, we use the explicit kernel formula (19) and the procedure outlined at the end of Section 5 to convert this representation into (11), which removes the artificial dependence upon $\boldsymbol{\phi}$. ∎

Proof of Theorem 2. From the proof of Proposition 8, we know that the minimal singular value of the cross-product matrix $\mathbf{P} = [\mathbf{p}_1 \cdots \mathbf{p}_M]^T$ with $\mathbf{p}_m = \boldsymbol{\nu}(p_m) \in \mathbb{R}^{N_0}$ is non-vanishing. The geometric implication is that $\mathrm{span}\{\mathbf{p}\}_{m=1}^M = \mathbb{R}^{N_0}$. Since the corresponding system is redundant, we can always identify a subset of these row vectors that forms a basis of $\mathbb{R}^{N_0}$. Without loss of generality, we now assume that this subset is $\{\mathbf{p}_m\}_{m=1}^{N_0}$ and that the corresponding submatrix $\mathbf{P}_0 = [\mathbf{p}_1 \cdots \mathbf{p}_{N_0}]^T$ is therefore invertible. In other words, we have identified a reduced vector of measurement functionals $\boldsymbol{\nu}_0 = (\nu_1, \ldots, \nu_{N_0})$ that is linearly independent with respect to $\mathcal{N}_{\mathrm{L}}$. This, in turn, allows us to construct the boundary functional $\boldsymbol{\phi}_0 = \mathbf{P}_0^{-1} \boldsymbol{\nu}_0$ that meets the biorthogonal requirement $\boldsymbol{\phi}_0(p_n) = \mathbf{P}_0^{-1} \boldsymbol{\nu}_0(p_n) = \mathbf{P}_0^{-1} \mathbf{p}_n = \mathbf{e}_n$. In effect, this yields a biorthogonal system with the property that $\mathcal{N}_{\mathrm{L}}' = \mathrm{span}\{\nu_n\}_{n=1}^{N_0} \subset \mathcal{M}_{\mathrm{L}}'(\mathbb{R}^d)$.

Coming back to our interpolation problem, we define $\mathbf{y} = (\mathbf{y}_0, \mathbf{y}_1)$ with $\mathbf{y}_0 = (y_1, \ldots, y_{N_0}) \in \mathbb{R}^{N_0}$ and $\mathbf{y}_1 \in \mathbb{R}^{M-N_0}$. Due to the biorthogonality of $(\boldsymbol{\phi}_0, \boldsymbol{p})$, the unique element $q_0 \in \mathcal{N}_{\mathrm{L}}$ such that $\boldsymbol{\nu}_0(q_0) = \mathbf{P}_0 \boldsymbol{\phi}_0(q_0) = \mathbf{y}_0$ is given by

$$q_0 = \sum_{n=1}^{N_0} b_n p_n$$

with $\mathbf{b} = (b_1, \ldots, b_{N_0}) = \mathbf{P}_0^{-1} \mathbf{y}_0$. The other ingredient is Theorem 5, which ensures that any $f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ has a unique decomposition as $f = \mathrm{L}_{\boldsymbol{\phi}_0}^{-1} w + q$ with $q \in \mathcal{N}_{\mathrm{L}}$ and $w \in \mathcal{M}(\mathbb{R}^d)$. Now, the crucial property is that the boundary conditions $\boldsymbol{\phi}_0(\mathrm{L}_{\boldsymbol{\phi}_0}^{-1} w) = \mathbf{0}$ imply that $\boldsymbol{\nu}_0(\mathrm{L}_{\boldsymbol{\phi}_0}^{-1} w) = \mathbf{0}$ for all $w \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$. This allows us rewrite the solution of our generalized interpolation problem as $f = \mathrm{L}_{\boldsymbol{\phi}_0}^{-1} w_1 + q_0$, where

$$w_1 = \arg \min_{w \in \mathcal{M}(\mathbb{R}^d)} \|w\|_{\mathcal{M}} \quad \text{s.t.} \quad \boldsymbol{\nu}_1(\mathrm{L}_{\boldsymbol{\phi}_0}^{-1} w) = \mathbf{y}_1 - \boldsymbol{\nu}_1(q_0).$$

The result then follows from the continuity of $\mathrm{L}_{\boldsymbol{\phi}_0}^{-1}$ and the reduced version of Theorem 7 with $\mathcal{N} = \{0\}$. ∎

8. Further Theoretical and Computational Issues. The analogy with the finite-dimensional theory of compressed sensing raises the fundamental theoretical question: Is it possible to provide conditions on the measurement operator $\boldsymbol{\nu}$ such that a perfect recovery is possible for certain classes of signals; in particular, splines with a given number of knots? This is an open topic that calls for further investigation. Because the problem is formulated in the continuum, we suspect that it is much more difficult—if not impossible—to identify conditions that ensure unicity.

While the reconstruction problem in Theorem 1 is formulated in *analysis form* (*i.e.*, minimization of $\|\mathrm{L}s\|_{\mathcal{M}}$), the interesting outcome is that the solution (11) is given in *synthesis form*, with the unusual twist that the underlying dictionary $\{\rho_{\mathrm{L}}(\cdot - \boldsymbol{\tau})\}_{\boldsymbol{\tau} \in \mathbb{R}^d}$ of basis functions is infinite-dimensional and not even countable. This interpretation suggests a natural discretization which is to select a finite subset of equally-spaced functions $\{\rho_{\mathrm{L}}(\cdot - \boldsymbol{\tau}_n)\}_{n=1}^N$ with $N \gg M$ and to rely on linear programming for $\epsilon = 0$, or quadratic programming for $\epsilon > 0$, or some other convex optimization technique to numerically solve the underlying ℓ_1-minimization problem.

We have preliminary evidence that this approach is feasible. In particular, we have considered the generalized interpolation scenario covered by Theorem 2 and observed that the simplex algorithm performs well in the sense that it always returns a nonuniform L-spline with a number of knots $K \leq (M - N_0)$. The key theoretical question now is to establish the convergence of such a scheme as the sampling step gets smaller.

Since the space that is spanned by the null-space components of L and the integer shifts of ρ_{L} is the space of cardinal L-splines (see [52] for the generic case of an ordinary differential operator), one may also consider an alternative discretization that uses the corresponding B-spline basis functions, which are much better conditioned than Green's functions. This would bring us back to a numerical problem that is very similar to (3), with the advantage of maintaining a direct control over the discretization error. In the case of a pure denoising problem, another possible option is to run the taut-string algorithm [38, 44] or some appropriate variation thereof.

At any rate, we believe that the issue of the proper discretization of the reconstruction problem (12) as well as the development of adequate numerical schemes are important research topics on their own right. For the cases where the solution is not unique, Theorem 1 also suggests a new computational challenge: the design of a minimization algorithm that systematically converges to an extremal point of the problem, the best solution being the spline that exhibits the minimal number of knots $K = \|\mathbf{a}\|_0$.

Appendix A. Proof of Theorem 3.

Proof. First, we establish the sufficiency of the stability condition by considering the signal $f(\boldsymbol{x}) = \mathrm{G}\{w\}(\boldsymbol{x}) = \int_{\mathbb{R}^d} g(\boldsymbol{x}, \boldsymbol{y}) w(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}$, where $w \in \mathcal{M}(\mathbb{R}^d)$, and by constructing the estimate

$$\begin{aligned} |f(\boldsymbol{x})|(1 + \|\boldsymbol{x}\|)^{-\alpha_0} &= (1 + \|\boldsymbol{x}\|)^{-\alpha_0} \left| \int_{\mathbb{R}^d} g(\boldsymbol{x}, \boldsymbol{y}) w(\boldsymbol{y}) \mathrm{d}\boldsymbol{y} \right| \\ &\leq (1 + \|\boldsymbol{x}\|)^{-\alpha_0} \operatorname*{ess\,sup}_{\boldsymbol{y} \in \mathbb{R}^d} |g(\boldsymbol{x}, \boldsymbol{y})| \, \|w\|_{\mathcal{M}}, \end{aligned}$$

which implies that

$$\|f\|_{\infty,\alpha_0} = \|\mathrm{G}\{w\}\|_{\infty,\alpha_0} \leq \left(\operatorname*{ess\,sup}_{\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d} |g(\boldsymbol{x}, \boldsymbol{y})| \, (1 + \|\boldsymbol{x}\|)^{-\alpha_0} \right) \|w\|_{\mathcal{M}}$$

for all $w \in \mathcal{M}(\mathbb{R}^d)$. In doing so, we have shown that

$$\|\mathrm{G}\| \leq \operatorname*{ess\,sup}_{\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d} |g(\boldsymbol{x}, \boldsymbol{y})| \, (1 + \|\boldsymbol{x}\|)^{-\alpha_0} < \infty.$$

To prove necessity, we use the property that $g(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{G}\{\delta(\cdot - \boldsymbol{y})\}(\boldsymbol{x})$, where the shifted Dirac impulse $\delta(\cdot - \boldsymbol{y})$ is included in $\mathcal{M}(\mathbb{R}^d)$ with $\|\delta(\cdot - \boldsymbol{y})\|_{\mathcal{M}} = 1$. We then observe that, for each $\boldsymbol{y} \in \mathbb{R}^d$,

$$\|\mathrm{G}\{\delta(\cdot - \boldsymbol{y})\}\|_{\infty,\alpha_0} = \operatorname*{ess\,sup}_{\boldsymbol{x} \in \mathbb{R}^d} (1 + \|\boldsymbol{x}\|)^{-\alpha_0} |g(\boldsymbol{x}, \boldsymbol{y})|.$$

Moreover, G being bounded, we have that

$$\|\mathrm{G}\{\delta(\cdot - \boldsymbol{y})\}\|_{\infty,\alpha_0} \leq \|\delta(\cdot - \boldsymbol{y}) \, \|_{\mathcal{M}} \, \|G\| = \|G\|,$$

which means that

$$\operatorname*{ess\,sup}_{\boldsymbol{x}\in\mathbb{R}^d}(1+\|\boldsymbol{x}\|)^{-\alpha_0}|g(\boldsymbol{x},\boldsymbol{y})| \leq \operatorname*{ess\,sup}_{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d}(1+\|\boldsymbol{x}\|)^{-\alpha_0}|g(\boldsymbol{x},\boldsymbol{y})| \leq \|G\| < \infty.$$

As we already know that the inequality holds in the other direction as well, we obtain

$$\|\mathrm{G}\| = \operatorname*{ess\,sup}_{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d}|g(\boldsymbol{x},\boldsymbol{y})|\ (1+\|\boldsymbol{x}\|)^{-\alpha_0},$$

which concludes the proof. ∎

Appendix B. Proof of Theorem 4.

Proof. We first establish the properties of the operator on Schwartz' space of smooth and rapidly-decreasing signals $\mathcal{S}(\mathbb{R}^d)$ to avoid any technical problem related to splitting sums and interchanging integrals. We also rely on Schwartz's kernel theorem which states the equivalence between the continuous linear operators $\mathrm{G}:\mathcal{S}(\mathbb{R}^d)\to\mathcal{S}'(\mathbb{R}^d)$ and their Schwartz kernels (or general impulse response) $g\in\mathcal{S}'(\mathbb{R}^d\times\mathbb{R}^d)$, meaning that two such operators are identical if and only if their kernels are equal—in the sense of distributions.

Using the explicit representation of $g_{\boldsymbol{\phi}}$ together with $\mathrm{L}\{\rho_{\mathrm{L}}\}=\delta$ (Dirac distribution) and $\mathrm{L}\{p_m\}=0$ for $m=1,\dots,N_0$, we then easily show that

$$\mathrm{L}\mathrm{L}_{\boldsymbol{\phi}}^{-1}\{\varphi\}=\varphi$$

for all $\varphi\in\mathcal{S}(\mathbb{R}^d)$. Next, we invoke the biorthogonality property $\langle\phi_m,p_n\rangle=\delta_{m-n}$ (Kronecker delta) to evaluate the inner product of (19) with ϕ_m as

$$\begin{aligned}\langle\phi_m,\mathrm{L}_{\boldsymbol{\phi}}^{-1}\{\varphi\}\rangle &= \langle\phi_m,\rho_{\mathrm{L}}*\varphi\rangle-\sum_{n=1}^{N_0}\langle\phi_m,p_n\rangle\langle q_n,\varphi\rangle\\ &= \langle\phi_m,\rho_{\mathrm{L}}*\varphi\rangle-\langle q_m,\varphi\rangle\\ &= \langle\phi_m,\rho_{\mathrm{L}}*\varphi\rangle-\langle\phi_m,\rho_{\mathrm{L}}*\varphi\rangle=0,\end{aligned}$$

which proves that the boundary conditions are satisfied.

Let us now consider another operator L^{-1} that is also a right inverse of L. Clearly, the result of the action of the two operators can only differ by a component that is in the null space of L so that $(\mathrm{L}^{-1}\varphi-\mathrm{L}_{\boldsymbol{\phi}}^{-1}\varphi)=q\in\mathcal{N}_{\mathrm{L}}$. Since $(\boldsymbol{\phi},\boldsymbol{p})$ forms a biorthogonal system, q is uniquely determined by $\boldsymbol{\phi}(q)$, which implies that the right-inverse operator that imposes the condition $\boldsymbol{\phi}(\mathrm{L}^{-1}\varphi)=\mathbf{0}$ is unique.

Next, we define $C=\sup_{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d}\left(|g_{\boldsymbol{\phi}}(\boldsymbol{x},\boldsymbol{y})|\ (1+\|\boldsymbol{x}\|)^{-n_0}\right)<\infty$ and extract the generic continuity bound

$$\|\mathrm{L}_{\boldsymbol{\phi}}^{-1}\varphi\|_{\infty,n_0}\leq C\|\varphi\|_{\mathcal{M}}$$

from the the proof of Theorem 3. This allows us to extend the domain of the operator from $\mathcal{S}(\mathbb{R}^d)$ to $\mathcal{M}(\mathbb{R}^d)$, by the Hahn-Banach theorem. Since $\mathcal{S}(\mathbb{R}^d)$ is dense in $\mathcal{M}(\mathbb{R}^d)$, we can do likewise for the right-inverse property and the boundary conditions by invoking the continuity of $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ and $\boldsymbol{\phi}$.

Alternatively, one can establish this extension indirectly by identifying a specific Banach space $\mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d)$ and then by showing that it is the bijective image of $\mathcal{M}(\mathbb{R}^d)$ by $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ (see proof of Statement 1 in Theorem 5, which also nicely settles the issue of stability). ∎

Appendix C. Proof of Theorem 6.

Proof. First, we prove that $C_{\mathrm{L},\boldsymbol{p}}(\mathbb{R}^d)$ is isometrically isomorphic to $C_0(\mathbb{R}^d)$. For any $\varphi \in C_0(\mathbb{R}^d)$, $\mathrm{L}^*\varphi = 0$ implies that $\varphi \in \mathcal{N}_{\mathrm{L}^*} \cap C_0(\mathbb{R}^d) = \{0\}$ (since the basis functions of the null space do not vanish at infinity); *i.e.*, $\varphi = 0$. L^* is therefore injective, and hence bijective since it is surjective $C_0(\mathbb{R}^d) \to C_{\mathrm{L},\boldsymbol{p}}(\mathbb{R}^d)$ by definition. In particular, this implies that the adjoint $\mathrm{L}_{\boldsymbol{\phi}}^{-1*}$ of the operator $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ defined by (19) is the inverse of L^* from $C_{\mathrm{L},\boldsymbol{p}}(\mathbb{R}^d)$ to $C_0(\mathbb{R}^d)$. Therefore, $C_{\mathrm{L},\boldsymbol{p}}(\mathbb{R}^d)$ inherits the Banach-space structure of $C_0(\mathbb{R}^d)$ for the norm $\|\mathrm{L}_{\boldsymbol{\phi}}^{-1} f\|_\infty$.

If $f \in C_{\mathrm{L},\boldsymbol{p}}(\mathbb{R}^d)$, then $f = \mathrm{L}^*\varphi$ with $\varphi \in C_0(\mathbb{R}^d)$ and $\langle f, p\rangle = \langle \varphi, \mathrm{L}p\rangle = 0$ for any $p \in \mathcal{N}_{\mathrm{L}}$. The unique element of $\mathcal{N}_{\mathrm{L}}'$ orthogonal to $\mathcal{N}_{\mathrm{L}}$ is 0 so that the sum $C_{\mathrm{L},\boldsymbol{p}}(\mathbb{R}^d) \oplus \mathcal{N}_{\mathrm{L}}'$ is direct. $\mathcal{N}_{\mathrm{L}}'$ is a (finite-dimensional) Banach space for the norm $\|\boldsymbol{p}(f)\|_2$, implying that $C_{\mathrm{L}}(\mathbb{R}^d) = C_{\mathrm{L},\boldsymbol{p}}(\mathbb{R}^d) \oplus \mathcal{N}_{\mathrm{L}}'$ is a Banach space for (24).

Next, we recall that $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ is continuous and bijective from $\mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d)$ to $\mathcal{M}(\mathbb{R}^d)$ (Theorem 5), while we have just shown that its adjoint is continuous and bijective from $C_{\mathrm{L},\boldsymbol{p}}(\mathbb{R}^d)$ to $C_0(\mathbb{R}^d)$. Knowing that $\big(C_0(\mathbb{R}^d)\big)' = \mathcal{M}(\mathbb{R}^d)$, this implies that $\big(C_{\mathrm{L},\boldsymbol{p}}(\mathbb{R}^d)\big)' = \mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d)$. Finally, we have

$$(C_{\mathrm{L}}(\mathbb{R}^d))' = (C_{\mathrm{L},\boldsymbol{p}}(\mathbb{R}^d) \oplus \mathcal{N}_{\mathrm{L}}')' = (C_{\mathrm{L},\boldsymbol{p}}(\mathbb{R}^d))' \oplus (\mathcal{N}_{\mathrm{L}}')' = \mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d) \oplus \mathcal{N}_{\mathrm{L}} = \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d),$$

as expected.

To establish the weighted L_1-norm inequality, we first observe that the hypotheses $f \in L_{1,-n_0}(\mathbb{R}^d)$ and $p_n \in L_{\infty,n_0}(\mathbb{R}^d)$ imply that $|\langle f, p_n\rangle| \leq \|p_n\|_{\infty,n_0} \|f\|_{L_{1,-n_0}}$ (by the Hölder inequality). Likewise, using the stability bound (21), we get

$$\begin{aligned} |\mathrm{L}_{\boldsymbol{\phi}}^{-1*}\{f\}(\boldsymbol{y})| &= \left| \int_{\mathbb{R}^d} g_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{y})\, f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \right| \\ &\leq \int_{\mathbb{R}^d} C_{\boldsymbol{\phi}} (1 + \|\boldsymbol{x}\|)^{n_0} |f(\boldsymbol{x})| \mathrm{d}\boldsymbol{x} = C_{\boldsymbol{\phi}} \|f\|_{L_{1,-n_1}}, \end{aligned}$$

which yields $\|\mathrm{L}_{\boldsymbol{\phi}}^{-1*}\{f\}\|_\infty \leq C_{\boldsymbol{\phi}} \|f\|_{L_{1,-n_1}}$. The desired result is then obtained from the summation of these individual bounds. ∎

Appendix D. Proof of Theorem 7.

The proof follows the same steps as the original one of Fisher and Jerome [27, Theorem 1]. Yet, it differs in the assumptions and technicalities (*i.e.*, the consideration of the non-compact domain $\mathbb{R}^d$ and the use of explicit bounds). We have done our best to make it self-contained.

As preparation, we recall that the weak*-topology on $\mathcal{M}(\mathbb{R}^d) = \big(\mathcal{C}_0(\mathbb{R}^d)\big)'$ is the locally convex topology associated with the family of semi-norms $p_\varphi(w) = |\langle w, \varphi\rangle|$ for $\varphi \in C_0(\mathbb{R}^d)$. In particular, a sequence of elements $w_n \in \mathcal{M}(\mathbb{R}^d)$ converges to 0 for the weak*-topology if and only if $\langle w_n, \varphi\rangle \to 0$ for every $\varphi \in C_0(\mathbb{R}^d)$. A subset of $\mathcal{M}(\mathbb{R}^d)$ is said to be weak*-closed (weak*-compact, respectively) if it is closed (compact, respectively) for the weak*-topology. We shall use Propositions 9 and 10, which are consequences of the Banach-Alaoglu theorem and its variations [43, p.68].

Proposition 9. *Compactness in the weak*-topology of $\mathcal{M}(\mathbb{R}^d)$.*

- *For every $\alpha > 0$, the set $\mathcal{B}_\alpha = \{w \in \mathcal{M}(\mathbb{R}^d) : \|w\|_{\mathcal{M}} \leq \alpha\}$ is weak*-compact in $\mathcal{M}(\mathbb{R}^d)$.*
- *If (w_n) is a sequence in $\mathcal{M}(\mathbb{R}^d)$, bounded for the TV-norm, then we can extract a subsequence that converges in $\mathcal{M}(\mathbb{R}^d)$ for the weak*-topology.*

The second point of Proposition 9 is valid because the space $C_0(\mathbb{R}^d)$ is separable. These properties also carry over to the Banach space $\mathcal{H} = \mathcal{M}(\mathbb{R}^d) \oplus \mathcal{N} = \big(C_0(\mathbb{R}^d) \oplus \mathcal{N}'\big)'$, which is endowed with the corresponding weak*-topology: A sequence (w_n, p_n) in $\mathcal{H}$ vanishes for the weak*-topology if and only if w_n vanishes for the weak*-topology of $\mathcal{M}(\mathbb{R}^d)$ and $\|p_n\|_{\mathcal{N}} \to 0$.

PROPOSITION 10. *Compactness in the weak*-topology of $\mathcal{H}$.*

- *For every $\alpha_1, \alpha_2 > 0$, the set $\mathcal{B}_{\alpha_1,\alpha_2} = \{(w,p) \in \mathcal{H} :\ \|w\|_{\mathcal{M}} \leq \alpha_1,\ \|p\|_{\mathcal{N}} \leq \alpha_2\}$ is weak*-compact in $\mathcal{H}$.*
- *If (w_n, p_n) is a sequence in $\mathcal{H}$ such that $\|w_n\|_{\mathcal{M}} + \|p_n\|_{\mathcal{N}}$ is bounded, then we can extract a subsequence that converges in $\mathcal{H}$ for the weak*-topology.*

Proof of Theorem 7. The proof is divided in two parts. First, we show that $\mathcal{V}$ is a nonempty, convex, and weak*-compact subspace of $\mathcal{H}$. This allows us to specify $\mathcal{V}$ by means of its extremal points via the Krein-Milman theorem. Second, we show that the extremal points have the announced form. We set $\beta = \inf_{(w,p)\in\mathcal{U}} \|w\|_{\mathcal{M}}$.

Part I: $\mathcal{V}$ is nonempty, convex, and weak-compact.* Since $\boldsymbol{F}$ is weak*-continuous, it is also continuous $\mathcal{H} \to \mathbb{R}^M$ in the topology of $\mathcal{H}$. Hence, there exists a constant $A > 0$ such that

$$\|\boldsymbol{F}(w,p)\|_2 \leq A(\|w\|_{\mathcal{M}} + \|p\|_{\mathcal{N}_{\mathrm{L}}}). \tag{30}$$

Let us consider a sequence $(w_n, p_n)_{n\in\mathbb{N}}$ in $\mathcal{U}$ such that $\|w_n\|_{\mathcal{M}}$ decreases to β. In particular, $\|w_n\|_{\mathcal{M}}$ is bounded above by $\|w_0\|_{\mathcal{M}}$. We set $M = \max_{\boldsymbol{x}\in\mathcal{C}} \|\boldsymbol{x}\|$. Using respectively (26), (30), and $\|\boldsymbol{F}(w_n, p_n)\|_2 \leq M$ (since $(w_n, p_n) \in \mathcal{U}$), we deduce the inequalities

$$\begin{aligned} \|p_n\|_{\mathcal{N}} &\leq \frac{1}{B}\|\boldsymbol{F}(0,p_n)\|_2 = \frac{1}{B}\|\boldsymbol{F}(w_n,p_n) - \boldsymbol{F}(w_n,0)\|_2 \\ &\leq \frac{1}{B}(\|\boldsymbol{F}(w_n,p_n)\|_2 + \|\boldsymbol{F}(w_n,0)\|_2) \\ &\leq \frac{1}{B}(M + A\|w_n\|_{\mathcal{M}}) \leq \frac{1}{B}(M + A\|w_0\|_{\mathcal{M}}), \end{aligned} \tag{31}$$

which shows that p_n is bounded. We can then extract a sequence (w_{s_n}, p_{s_n}) from (w_n, p_n) that converges to $(w_\infty, p_\infty) \in \mathcal{H}$ for the weak*-topology (by Proposition 10). Since $\|w_n\| \to \beta$ and $\|w_{s_n}\|$ is a subsequence, it must also converge to β.

On the other hand, the set $\mathcal{U} = \boldsymbol{F}^{-1}(\mathcal{C})$ is weak*-closed in $\mathcal{H}$, as the preimage of a closed set by a weak*-continuous function $\boldsymbol{F}$. Consequently, (w_∞, p_∞) is the weak*-limit of a sequence of elements in $\mathcal{U}$. We therefore deduce that $(w_\infty, p_\infty) \in \mathcal{U}$, so that $\|w_\infty\|_{\mathcal{M}} \geq \beta$. In light of the previous inequality, this yields $\|w_\infty\|_{\mathcal{M}} = \beta$, which proves that $\mathcal{V}$ is not empty.

In addition to being weak*-closed, the set $\mathcal{U} = \boldsymbol{F}^{-1}(\mathcal{C})$ is convex because $\mathcal{C}$ is convex and $\boldsymbol{F}$ linear. Likewise, $\mathcal{V} = \mathcal{U} \bigcap \{(w,p) :\ \|w\|_{\mathcal{M}} \leq \beta\}$ is convex, weak*-closed as the intersection of two sets with the same property. Finally, for $(w,p) \in \mathcal{V}$, we show that $\|p\|_{\mathcal{N}} \leq \frac{M+A\|w\|_{\mathcal{M}}}{B} = \frac{M+A\beta}{B} = \gamma$, based on the same inequalities as in (31). Therefore, we have

$$\mathcal{V} \subset \{(w,p) \in \mathcal{H} :\ \|w\|_{\mathcal{M}} \leq \beta,\ \|p\|_{\mathcal{N}} \leq \gamma\},$$

where the set on the right-hand side is weak*-compact, due to Proposition 10. Since any weak*-closed set included in a weak*-compact set is necessarily weak*-compact,

this shows that $\mathcal{V}$ is weak*-compact.

We are now in the position to apply the Krein-Milman theorem [43, p. 75] to the convex weak*-compact set $\mathcal{V} \subset \mathcal{H}$, which tells us that "$\mathcal{V}$ is the closed convex hull of its extreme points in $\mathcal{H}$ endowed with the weak*-topology". This leads us to the final part of the proof, which is the characterization of those extreme points.

Part II: The extreme points of $\mathcal{V}$ are of the form (27). We shall prove that a necessary condition for (w, p) to be an extreme point of $\mathcal{V}$ is that there are no disjoint Borelian sets $E_1, \dots, E_{M+1} \subset \mathbb{R}^d$ such that $w(E_m) \neq 0$ for $m = 1, \dots, M+1$. The only elements of $\mathcal{M}(\mathbb{R}^d)$ satisfying this condition are precisely those described by (27).

We shall proceed by contradiction and assume that there exist disjoint sets $E_1, \dots, E_{M+1}$ such that $w(E_m) \neq 0$ for all m.

We denote the restriction of w to E_m as $w_m = w\mathbb{1}_{E_m}$. We also define $E = \mathbb{R}^d \backslash \bigcup_m E_m$, and $\bar{w} = w\mathbb{1}_E$ with $\|w\|_{\mathcal{M}} = \beta$. For $m = 1, \dots, M+1$, we set $\boldsymbol{y}_m = \boldsymbol{F}(w_m, p) \in \mathbb{R}^M$. Since any collection of $(M+1)$ vectors in $\mathbb{R}^M$ is linearly dependent, there exists $(c_m)_{1 \leq m \leq M+1} \neq \mathbf{0}$ such that $\sum_{m=1}^{M+1} c_m \boldsymbol{y}_m = \mathbf{0}$.

Let $\mu = \sum_{m=1}^{M+1} c_m w_m \in \mathcal{M}(\mathbb{R}^d)$ and $\epsilon \in (-\epsilon_{\max}, \epsilon_{\max})$ with $\epsilon_{\max} = 1/\max_m |c_m|$, so that $(1 + \epsilon c_m) > 0$ and $(1 - \epsilon c_m) > 0$ for all m. By construction, we have that

$$\boldsymbol{F}(\mu, p) = \sum_{m=1}^{M+1} c_m \boldsymbol{F}(w_m, p) = \sum_{m=1}^{M+1} c_m \boldsymbol{y}_m = \mathbf{0} \tag{32}$$

and, therefore, that $\boldsymbol{F}(w \pm \epsilon\mu, p) = \boldsymbol{F}(w, p)$. Hence,

$$(w \pm \epsilon\mu, p) \in \mathcal{U}.$$

Moreover, $w \pm \epsilon\mu = \bar{w} + \sum_{m=1}^{M+1} (1 \pm \epsilon c_m) w_m$. Since the measures $\bar{w}, w_1, \dots, w_{M+1}$ have disjoint supports and $(1 \pm \epsilon c_m) > 0$, we have

$$\begin{aligned} \|w \pm \epsilon\mu\|_{\mathcal{M}} &= \|\bar{w}\|_{\mathcal{M}} + \sum_{m=1}^{M+1} (1 \pm \epsilon c_m) \|w_m\|_{\mathcal{M}} \\ &= \|\bar{w}\|_{\mathcal{M}} + \sum_{m=1}^{M+1} \|w_m\|_{\mathcal{M}} \pm \epsilon \sum_{m=1}^{M+1} c_m \|w_m\|_{\mathcal{M}} \\ &= \|w\|_{\mathcal{M}} \pm \epsilon \sum_{m=1}^{M+1} c_m \|w_m\|_{\mathcal{M}} \\ &= \beta \pm \epsilon \sum_{m=1}^{M+1} c_m \|w_m\|_{\mathcal{M}}. \end{aligned} \tag{33}$$

If $\sum_{m=1}^{M+1} c_m \|w_m\|_{\mathcal{M}} \neq 0$, then we either have $\|w + \epsilon\mu\|_{\mathcal{M}} < \beta$ or $\|w - \epsilon\mu\|_{\mathcal{M}} < \beta$, which is impossible since the minimum over $\mathcal{U}$ is β. Hence,

$$\sum_{m=1}^{M+1} c_m \|w_m\|_{\mathcal{M}} = 0$$

and $\|w + \epsilon\mu\|_{\mathcal{M}} = \|w - \epsilon\mu\|_{\mathcal{M}} = \beta$, which translates into $(w + \epsilon\mu, p)$ and $(w - \epsilon\mu, p)$ being included in $\mathcal{V}$. This, in turn, implies that $(w, p) = \frac{1}{2}(w + \epsilon\mu, p) + \frac{1}{2}(w - \epsilon\mu, p)$ is not an extreme point of $\mathcal{V}$. . ∎

Acknowledgments. The research was partially supported by the Swiss National Science Foundation under Grant 200020-162343, the Center for Biomedical Imaging (CIBM) of the Geneva-Lausanne Universities and EPFL, and the European Research Council under Grant 692726 (H2020-ERC Project GlobalBioIm). The authors are thankful to H. Gupta for helpful discussions. They are also grateful to the three anonymous reviewers for their thoughtful suggestions which have helped improve the manuscript.

REFERENCES

[1] B. Adcock and A. Hansen, *Generalized sampling and infinite-dimensional compressed sensing*, Foundations of Computional Mathematics, (2015), pp. 1–61.
[2] B. Adcock, A. C. Hansen, C. Poon, and B. Roman, *Breaking the coherence barrier: A new theory for compressed sensing*, Forum of Mathematics, Sigma, (to appear).
[3] A. Beck and M. Teboulle, *Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems*, IEEE Transactions on Image Processing, 18 (2009), pp. 2419–2434.
[4] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 183–202.
[5] K. Bredies and H. Pikkarainen, *Inverse problems in spaces of measures*, ESAIM: Control, Optimisation and Calculus of Variations, 19 (2013), pp. 190–218.
[6] A. M. Bruckstein, D. L. Donoho, and M. Elad, *From sparse solutions of systems of equations to sparse modeling of signals and images*, SIAM Review, 51 (2009), pp. 34–81.
[7] E. J. Candès, *The restricted isometry property and its implications for compressed sensing*, Comptes Rendus de l'Académie des Sciences, 346 (2008), pp. 589–592.
[8] E. J. Candès and C. Fernandez-Granda, *Super-resolution from noisy data*, Journal of Fourier Analysis and Applications, 19 (2013), pp. 1229–1254.
[9] E. J. Candès and J. Romberg, *Sparsity and incoherence in compressive sampling*, Inverse Problems, 23 (2007), pp. 969–985.
[10] E. J. Candès, J. K. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Communications on Pure and Applied Mathematics, 59 (2006), pp. 1207–1223.
[11] A. Chambolle, *An algorithm for total variation minimization and applications*, Journal of Mathematical Imaging and Vision, 20 (2004), pp. 89–97.
[12] A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore, *Harmonic analysis of the space BV*, Revista Matematica Iberoamericana, 19 (2003), pp. 235–263.
[13] W. Dahmen and C. Micchelli, *On theory and application of exponential splines*, in Topics in Multivariate Approximation, C. Chui, L. Schumaker, and F. Utreras, eds., Academic Press, New York, 1987, pp. 37–46.
[14] I. Daubechies, M. Defrise, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Communications on Pure and Applied Mathematics, 57 (2004), pp. 1413–1457.
[15] C. de Boor, *A Practical Guide to Splines*, Springer-Verlag, New York, 1978.
[16] Q. Denoyelle, V. Duval, and G. Peyré, *Support recovery for sparse deconvolution of positive measures*, arXiv preprint arXiv:1506.08264, (2015).
[17] J. Deny and J.-L. Lions, *Les espaces du type de Beppo Levi*, Annales de l'Institut Fourier, 5 (1954), pp. 305–370.
[18] N. Dey, L. Blanc-Féraud, C. Zimmer, P. Roux, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia, *Richardson-Lucy algorithm with total variation regularization for 3D confocal microscope deconvolution*, Microscopy Research and Technique, 69 (2006), pp. 260–266.
[19] D. L. Donoho, *Compressed sensing*, IEEE Transactions on Information Theory, 52 (2006), pp. 1289–1306.
[20] D. L. Donoho and M. Elad, *Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization*, Proceedings of the National Academy of Sciences, 100 (2003), pp. 2197–2202.
[21] J. Duchon, *Splines minimizing rotation-invariant semi-norms in Sobolev spaces*, in Constructive Theory of Functions of Several Variables, W. Schempp and K. Zeller, eds., Springer-Verlag, Berlin, 1977, pp. 85–100.
[22] V. Duval and G. Peyré, *Exact support recovery for sparse spikes deconvolution*, Foundations of Computational Mathematics, 15 (2015), pp. 1315–1355.

[23] M. Elad, *Sparse and Redundant Representations. From Theory to Applications in Signal and Image Processing*, Springer, 2010.
[24] Y. Eldar, *Sampling Theory: Beyond Bandlimited Systems*, Cambridge University Press, 2015.
[25] C. Fernandez-Granda, *Super-resolution of point sources via convex programming*, Information and Inference, 5 (2016), pp. 251–303.
[26] M. A. T. Figueiredo and R. D. Nowak, *An EM algorithm for wavelet-based image restoration*, IEEE Transactions on Image Processing, 12 (2003), pp. 906–916.
[27] S. Fisher and J. Jerome, *Spline solutions to L^1 extremal problems in one and several variables*, Journal of Approximation Theory, 13 (1975), pp. 73–83.
[28] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Springer, 2013.
[29] I. M. Gelfand and G. Shilov, *Generalized Functions. Vol. 1. Properties and Operations*, Academic Press, New York, USA, 1964.
[30] T. Goldstein and S. Osher, *The split Bregman method for L_1-regularized problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 323–343.
[31] L. Grafakos, *Classical and Modern Fourier Analysis*, Prentice Hall, 2004.
[32] Y. Hel-Or and P. C. Teo, *Canonical decomposition of steerable functions*, Journal of Mathematical Imaging and Vision, 9 (1998), pp. 83–95.
[33] L. Hörmander, *The Analysis of Linear Partial Differential Operators. I. Distribution Theory and Fourier Analysis*, vol. 256, Springer-Verlag, Berlin, 2nd ed., 1990.
[34] T. Kurokawa, *Riesz potentials, higher Riesz transforms and Beppo Levi spaces*, Hiroshima Math. J, 18 (1988), pp. 541–597.
[35] J. E. Lavery, *Shape-preserving, multiscale fitting of univariate data by cubic L_1 smoothing splines*, Computer Aided Geometric Design, 17 (2000), pp. 715–727.
[36] M. Lustig, D. L. Donoho, and J. M. Pauly, *Sparse MRI: The application of compressed sensing for rapid MR imaging*, Magnetic Resonance in Medicine, 58 (2007), pp. 1182–1195.
[37] W. R. Madych and S. A. Nelson, *Polyharmonic cardinal splines*, Journal of Approximation Theory, 60 (1990), pp. 141–156.
[38] E. Mammen and S. van de Geer, *Locally adaptive regression splines*, Annals of Statistics, 25 (1997), pp. 387–413.
[39] S. Ramani and J. Fessler, *Parallel MR image reconstruction using augmented Lagrangian methods*, IEEE Transactions on Medical Imaging, 30 (2011), pp. 694 –706.
[40] H. Rauhut, K. Schnass, and P. Vandergheynst, *Compressed sensing and redundant dictionaries*, IEEE Transactions on Information Theory, 54 (2008), pp. 2210–2219.
[41] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.
[42] W. Rudin, *Real and Complex Analysis*, McGraw-Hill, New York, 3rd ed., 1987.
[43] W. Rudin, *Functional Analysis*, McGraw-Hill, New York, 1991.
[44] O. Scherzer, *Taut-string algorithm and regularization programs with g-norm data fit*, Journal of Mathematical Imaging and Vision, 23 (2005), pp. 135–143.
[45] I. J. Schoenberg, *Contributions to the problem of approximation of equidistant data by analytic functions*, Quarterly of Applied Mathematics, 4 (1946), pp. 45–99, 112–141.
[46] M. H. Schultz and R. S. Varga, *L-splines*, Numerische Mathematik, 10 (1967), pp. 345–369.
[47] L. L. Schumaker, *Spline Functions: Basic Theory*, Cambridge University Pressity Press, Cambridge, 3rd edition ed., 2007.
[48] G. Steidl, S. Didas, and J. Neumann, *Splines in higher order TV regularization*, International Journal of Computer Vision, 70 (2006), pp. 241–255.
[49] R. Tibshirani, *Regression shrinkage and selection via the Lasso*, Journal of the Royal Statistical Society. Series B, 58 (1996), pp. 265–288.
[50] M. Unser, *Sampling—50 years after Shannon*, Proceedings of the IEEE, 88 (2000), pp. 569–587.
[51] M. Unser and T. Blu, *Fractional splines and wavelets*, SIAM Review, 42 (2000), pp. 43–67.
[52] M. Unser and T. Blu, *Cardinal exponential splines: Part I—Theory and filtering algorithms*, IEEE Transactions on Signal Processing, 53 (2005), pp. 1425–1449.
[53] M. Unser and T. Blu, *Self-similarity: Part I—Splines and operators*, IEEE Transactions on Signal Processing, 55 (2007), pp. 1352–1363.
[54] M. Unser, J. Fageot, and H. Gupta, *Representer theorems for sparsity-promoting ℓ_1 regularization*, IEEE Transactions on Information Theory, 62 (2016), pp. 5167–5180.
[55] M. Unser and P. D. Tafti, *An Introduction to Sparse Stochastic Processes*, Cambridge University Press, 2014.
[56] M. Vetterli, P. Marziliano, and T. Blu, *Sampling signals with finite rate of innovation*,

IEEE Transactions on Signal Processing, 50 (2002), pp. 1417–1428.
[57] J. Ward and M. Unser, *Approximation properties of Sobolev splines and the construction of compactly supported equivalents*, SIAM Journal on Mathematical Analysis, 46 (2014), pp. 1843–1858.
[58] H. Wendland, *Scattered Data Approximations*, Cambridge University Press, 2005.