Continuous-Domain Signal Reconstruction Using L_p-Norm Regularization

Pakshal Bohra and Michael Unser, *Fellow, IEEE*

***Abstract*—We focus on the generalized-interpolation problem. There, one reconstructs continuous-domain signals that honor discrete data constraints. This problem is infinite-dimensional and ill-posed. We make it well-posed by imposing that the solution balances data fidelity and some L_p-norm regularization. More specifically, we consider $p \geq 1$ and the multi-order derivative regularization operator $\mathrm{L} = \mathrm{D}^{N_0}$. We reformulate the regularized problem exactly as a finite-dimensional one by restricting the search space to a suitable space of polynomial splines with knots on a uniform grid. Our splines are represented in a B-spline basis, which results in a well-conditioned discretization. For a sufficiently fine grid, our search space contains functions that are arbitrarily close to the solution of the underlying problem where our constraint that the solution must live in a spline space would have been lifted. This remarkable property is due to the approximation power of splines. We use the alternating-direction method of multipliers along with a multiresolution strategy to compute our solution. We present numerical results for spatial and Fourier interpolation. Through our experiments, we investigate features induced by the L_p-norm regularization, namely, sparsity, regularity, and oscillatory behavior.**

***Index Terms*—Interpolation, regularization, L_p-norm, splines.**

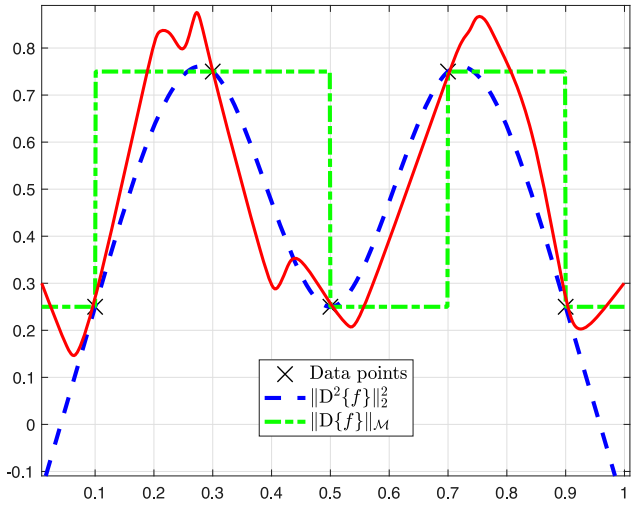

Fig. 1. Interpolation of data points symbolized by crosses. The solid line represents an arbitrary solution. For the other two cases, it is regularization that dictates how the points are connected.

I. Introduction

REGULARIZATION techniques are commonly used for the resolution of ill-posed problems. When these problems are formulated as optimization tasks, a standard way of applying regularization is to introduce a penalty term in the cost functional, which penalizes solutions with undesirable properties. For example, ℓ_2-regularization [1], [2] and, more recently, ℓ_p-regularization [3]–[7] has been widely used to reconstruct discrete-domain signals from their measurements.

In this work, we focus on problems where the object of interest f is defined over a continuum. In such cases, a natural candidate for the regularization term is $\|\mathrm{L}\{f\}\|$, where the choice of the operator L and the norm $\|\cdot\|$ allows us to incorporate prior knowledge about f. Continuous-domain regularization schemes such as Tikhonov [1], [8], [9], which uses the L_2-norm $\|\cdot\|_{L_2}$, and generalized total variation (gTV) [10], which involves the use of the $\mathcal{M}$-norm $\|\cdot\|_{\mathcal{M}}$ (an extension of the L_1-norm), have been intensively studied and their behavior is well-documented. To see the effect of these schemes, we consider the interpolation problem shown in Fig. 1. The objective there is to construct a continuously defined function that passes through the given data points exactly. However, as shown in the figure, it is possible to construct infinitely many valid solutions. In this problem, we regularize the solution by imposing a minimum-norm requirement of the form $\|\mathrm{L}\{f\}\|$. This enables us to obtain solutions with certain desired properties. It is well-known that Tikhonov (or L_2) regularization tends to produce smooth solutions while gTV regularization promotes sparsity. These characteristics can be seen in Fig. 1. For example, when we impose gTV regularization with $\mathrm{L} = \mathrm{D}$ (the derivative operator), we obtain a piecewise-constant solution whose derivative is sparse.

The purpose of this paper is to study the effect of continuous-domain L_p-norm regularization for a general $p \geq 1$ and a multi-order derivative operator $\mathrm{L} = \mathrm{D}^{N_0}$. To that end, we consider the generalized interpolation problem with L_p-norm regularization. Generalized interpolation is an extension of interpolation. Specifically, given certain measurement functionals $(\nu_1, \ldots, \nu_M)$ and a value (or measurement) y_m corresponding to each functional, we aim at constructing a continuously defined function that explains the measurements exactly. We formulate

Manuscript received August 7, 2019; revised March 20, 2020; accepted July 16, 2020. Date of publication August 3, 2020; date of current version August 20, 2020. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Weiyu Xu. This work was funded by Swiss National Science Foundation under Grant 200020_184646/1. *(Corresponding author: Pakshal Bohra.)*

The authors are with the Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (e-mail: pakshal.bohra@epfl.ch; michael.unser@epfl.ch).

Digital Object Identifier 10.1109/TSP.2020.3013781

1053-587X © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 09,2020 at 15:06:39 UTC from IEEE Xplore. Restrictions apply.

this problem as

$$\min_{f} \|\mathrm{D}^{N_0}\{f\}\|_{L_p} \text{ s.t. } \langle \nu_m, f\rangle = y_m,\ m = 1, 2, \ldots, M, \quad (1)$$

where $\|\cdot\|_{L_p}$ denotes the L_p-norm.

A. Why Generalized Interpolation?

The reconstruction of a signal f from a finite number of linear measurements $\mathbf{y} \in \mathbb{R}^M$ is a standard problem. Its ill-posedness is counteracted by regularization. Since the measurements are usually noisy, it is often formulated as the unconstrained optimization task

$$S = \arg\min_{f \in \mathcal{X}} \left(E\left(\mathbf{y}, \boldsymbol{\nu}(f)\right) + \lambda R(f)\right), \quad (2)$$

where $\mathcal{X}$ is a suitable function space, the operator $\boldsymbol{\nu}: f \mapsto \boldsymbol{\nu}(f) = (\langle \nu_1, f\rangle, \ldots, \langle \nu_M, f\rangle)$ describes the measurement model, $E: \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$ is a data-fidelity term which forces the solution to be consistent with the given measurements and R is the regularization. It can be shown[1] that, if E is strictly convex and R is convex, then all the solutions $f^* \in S$ generate the same measurement vector $\mathbf{z}_0 = \boldsymbol{\nu}(f^*) \in \mathbb{R}^M$. This property allows us to characterize the solution set S as

$$S = \arg\min_{f \in \mathcal{X}} R(f) \text{ s.t. } \boldsymbol{\nu}(f) = \mathbf{z}_0. \quad (3)$$

By understanding the effect of the regularization term $R(f)$ in (3), we can understand its effect for a much broader class of problems such as (2).

B. Related Work

The L_p-regularized interpolation problem and its variants, with $p \geq 1$ and $\mathrm{L} = \mathrm{D}^{N_0}$, have been studied in [11]–[16] in the context of approximation theory and splines. These works are theoretical, for the most part. They discuss the existence of a solution, conditions of optimality, and provide the functional form of the N_0th derivative of the solution. A specific instance of minimizing the L_p-norm of the second derivative of polynomial spline interpolants has been looked at in [17] and [18]. To the best of our knowledge, however, there exists no work that numerically solves the general continuous-domain problem (1) and demonstrates the effect of L_p-norm regularization.

C. Contributions

In this paper, we propose an algorithm to compute the solution to (1). Through a series of experiments, we then identify some properties of L_p-norm regularization. Our work complements [19], [20], where the recovery of signals is defined over a continuum, too, but without consideration of continuous-domain regularization. Here is a list of our contributions.

- We discretize the continuous-domain problem (1) by using a basis that consists of shifted polynomial B-splines of degree N_0, with knots on a uniform grid. This basis leads to an exact discretization, thus transforming our continuous-domain problem into an equivalent finite-dimensional discrete one which can be solved by algorithms such as the alternating-direction method of multipliers (ADMM) [21].
- We implement a multiresolution algorithm that progressively decreases the grid size until a solution with the desired precision is obtained. This strategy relies on the theory of approximation. It dictates that, when the grid size is sufficiently small, the search space spanned by our B-spline basis contains functions that are arbitrarily close to the solution of the full continuous-domain problem where our constraint that the solution must live in a spline space would have been lifted.
- We present numerical results for measurement operators that correspond to interpolation in the spatial and Fourier domains. In these experiments, we show the existence of a continuum of solutions as p varies from ∞ to 1. We then examine properties of L_p-regularized solutions such as sparsity, regularity (smoothness) and, oscillatory behavior and overshoot, as well as the effect of N_0 on the solutions.

The paper is organized as follows: In Section II, we introduce the continuous-domain framework and discuss some existing theoretical results. We provide background information on polynomial splines in Section III. Section IV includes the details of our discretization scheme, along with a discussion on the approximation power of splines. We present the multiresolution algorithm in Section V and illustrate our numerical results in Section VI.

II. Generalized Interpolation

In this section, we define and discuss the key components of our framework: the measurement operator, the regularization operator, the regularization norms, and the search space for the optimization problem. We then briefly review theoretical results available for this problem.

A. Continuous-Domain Framework

In generalized interpolation, the aim is to construct a function $f: \mathbb{R} \to \mathbb{R}$ that explains the measurements $\mathbf{y} \in \mathbb{R}^M$, with

$$\boldsymbol{\nu}(f) = (\langle \nu_1, f\rangle, \ldots, \langle \nu_M, f\rangle) = \mathbf{y}, \quad (4)$$

where $\langle \nu_m, f\rangle$ represents the action of the linear functional $\nu_m: f \mapsto \langle \nu_m, f\rangle = \nu_m(f) \in \mathbb{R}$. When ν_m and f are ordinary functions defined over $\mathbb{R}$, the mth measurement is given by the Lebesgue integral $\langle \nu_m, f\rangle = \int_{\mathbb{R}} \nu_m(x) f(x)\,\mathrm{d}x$. In the pure interpolation problem, the measurement functionals are shifted Dirac distributions $\nu_m = \delta(\cdot - x_m)$, with the property that $\langle \delta(\cdot - x_m), f\rangle = f(x_m)$.

In order to specify the regularization operator L, we introduce the Schwartz space $\mathcal{S}(\mathbb{R})$ of smooth and rapidly decaying functions defined over $\mathbb{R}$. Its continuous dual is the space of tempered distributions, denoted by $\mathcal{S}'(\mathbb{R})$. In our framework, we focus on regularization operators of the form $\mathrm{L} = \mathrm{D}^{N_0}: \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$, where D is the derivative operator extended to $\mathcal{S}'(\mathbb{R})$ [22, Chapter 3] and $N_0 \geq 1$. The null space of the operator D^{N_0} is $\mathcal{N}_{\mathrm{D}^{N_0}} = \mathrm{span}\{p_n\}_{n=1}^{N_0}$, with $p_n(x) = x^{n-1}$. The Green's function of D^{N_0} is denoted by $\rho_{\mathrm{D}^{N_0}}$; it satisfies the property that $\mathrm{D}^{N_0}\{\rho_{\mathrm{D}^{N_0}}\} = \delta$. The Green's function is not unique due to the existence of the null space.

[1] See the Appendix.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 09,2020 at 15:06:39 UTC from IEEE Xplore. Restrictions apply.

Next, we specify the the continuous-domain L_p-norm. For a measurable function $w : \mathbb{R} \to \mathbb{R}$, the L_p-norm ($1 \leq p < \infty$) is defined as

$$\|w\|_{L_p} \triangleq \left(\int_{\mathbb{R}} |w(x)|^p \, \mathrm{d}x \right)^{\frac{1}{p}}, \tag{5}$$

while the L_∞-norm is defined as[2]

$$\|w\|_{L_\infty} \triangleq \operatorname*{ess\,sup}_{x \in \mathbb{R}} |w(x)|. \tag{6}$$

Equation (5) also specifies the L_p quasi-norm for values of $p \in (0, 1)$. The Lebesgue space of functions $L_p(\mathbb{R}) = \{w : \mathbb{R} \to \mathbb{R} \mid \|w\|_{L_p} < \infty\}$, where $p \in [1, \infty]$, is a Banach space. Here, we also define the $\mathcal{M}$-norm used in gTV regularization, which is closely related to L_1 regularization, as

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

for any $w \in \mathcal{S}'(\mathbb{R})$. The Banach space associated with $\|\cdot\|_{\mathcal{M}}$ is $\mathcal{M}(\mathbb{R}) = \{w \in \mathcal{S}'(\mathbb{R}) \mid \|w\|_{\mathcal{M}} < +\infty\}$. The $\mathcal{M}$-norm is an extension of the L_1-norm. Indeed, for any $w \in L_1(\mathbb{R})$, we have that

$$\|w\|_{\mathcal{M}} = \|w\|_{L_1}. \tag{8}$$

However, the Dirac impulse δ is included in $\mathcal{M}(\mathbb{R})$ with $\|\delta\|_{\mathcal{M}} = 1$ but does not belong to $L_1(\mathbb{R})$. Thus, we have that $L_1(\mathbb{R}) \subset \mathcal{M}(\mathbb{R})$.

Finally, we define the search spaces for the gTV-regularized and L_p-regularized problems as

$$\mathcal{M}^{(N_0)}(\mathbb{R}) = \{f \in \mathcal{S}'(\mathbb{R}) \mid \mathrm{D}^{N_0}\{f\} \in \mathcal{M}(\mathbb{R})\} \tag{9}$$

$$L_p^{(N_0)}(\mathbb{R}) = \{f \in \mathcal{S}'(\mathbb{R}) \mid \mathrm{D}^{N_0}\{f\} \in L_p(\mathbb{R})\}. \tag{10}$$

There, we consider all generalized functions in $\mathcal{S}'(\mathbb{R})$ for which the regularization term is finite.

Now that we have described all the components involved in our regularized generalized-interpolation framework, we state the optimization problems that we consider in this work. They are

$$\mathcal{S}_{\mathcal{M}} = \operatorname*{arg\,min}_{f \in \mathcal{M}^{(N_0)}(\mathbb{R})} \|\mathrm{D}^{N_0}\{f\}\|_{\mathcal{M}} \text{ s.t. } \boldsymbol{\nu}(f) = \mathbf{y} \tag{11}$$

$$\mathcal{S}_p = \operatorname*{arg\,min}_{f \in L_p^{(N_0)}(\mathbb{R})} \|\mathrm{D}^{N_0}\{f\}\|_{L_p} \text{ s.t. } \boldsymbol{\nu}(f) = \mathbf{y}, \tag{12}$$

where $N_0 \geq 1$.

B. Theoretical Results

Before the discussion of theoretical results, we need to make some assumptions.

Assumption 1: In the following statements, the symbol $\mathcal{X}$ represents the search space $\mathcal{M}^{(N_0)}(\mathbb{R})$ or $L_p^{(N_0)}(\mathbb{R})$, depending on the problem at hand.

[2]The essential supremum is a generalization of the supremum in Lebesgue's theory of integration. For a measurable function $w : \mathbb{R} \to \mathbb{R}$, it is the smallest value $a \in \mathbb{R}$ such that $w(x) \leq a$ almost everywhere (i.e., everywhere except on a set of measure zero). The essential supremum is equivalent to the supremum for continuous functions.

i) The measurement operator $\boldsymbol{\nu}$ is weak*-continuous on $\mathcal{X}$.
ii) For the given measurements $\mathbf{y} \in \mathbb{R}^M$ and measurement operator $\boldsymbol{\nu}$, there exists at least one function $f_0 \in \mathcal{X}$ such that $\boldsymbol{\nu}(f_0) = \mathbf{y}$.
iii) The intersection of the null spaces of $\boldsymbol{\nu}$ and D^{N_0} is $\{0\}$.

Assumption (1.i) implies that the measurement functionals satisfy $\nu_m \in \mathcal{Y}$ for $m = 1, \ldots, M$, where the predual space $\mathcal{Y}$ is such that $\mathcal{X} = \mathcal{Y}'$. In practice, this imposes a minimum degree of regularity and decay on $\{\nu_m\}_{m=1}^M$. Assumption (1.ii) states a feasibility condition and is needed to ensure that the generalized interpolation problem is well-defined. If (1.i) holds and the ν_m are linearly independent, then (1.ii) is satisfied for any $\mathbf{y} \in \mathbb{R}^M$. Assumption (1.iii) ensures that the problem is well-posed over the null space of the regularization operator, where the penalization is immaterial. This can be checked by verifying that the matrix $\mathbf{P}$ with entries $[\mathbf{P}]_{m,n} = \langle \nu_m, p_n \rangle$ is full-rank.

For the gTV-regularized and L_2-regularized problems, there exist representer theorems that provide a parametric characterization of the possible range of solutions. In the case of L_2 regularization, the solution is unique, smooth, and lies in a finite-dimensional subspace that depends on the measurement and regularization operators [9]. The gTV problem can have infinitely many solutions, but the extreme points of the solution set $\mathcal{S}_{\mathcal{M}}$ are known to be splines whose type depends on the regularization operator only [10]. These splines have adaptive knots which are fewer than the number of measurements. On applying the operator D^{N_0} to these extreme points, we recover Dirac impulses at the knot locations, which implies a sparse N_0th order derivative. We refer to such solutions as the sparse solutions of the gTV problem.

Beside providing insights about the nature of the solutions, the representer theorems also play a role in the design of numerical methods to solve these problems. The parametric forms of the solution provided by the theorems are used for the discretization of the continuous-domain problems, leading to finite-dimensional optimization tasks which can be solved using standard optimization algorithms. A detailed comparison of L_2 versus gTV regularization can be found in [9]. The reader can refer to [9], [23] for the algorithms.

In this work, our main focus is on (12) with a general $p \geq 1$. This kind of a problem has been addressed in [11] for the case of pure interpolation, when the measurement functionals are Dirac impulses. Here, we state the result from [11] in a form that is compatible with our framework. When $p \in (1, \infty)$, there exists a unique solution f_0 to the L_p-regularized interpolation problem. It satisfies

$$\mathrm{D}^{N_0}\{f_0\} = \frac{|v_0|^{q-1}}{\|v_0\|_{L_q}^{q-2}} \operatorname{sgn}(v_0), \tag{13}$$

where $\frac{1}{p} + \frac{1}{q} = 1$ and

$$v_0(x) = \sum_{m=1}^{M} a_m \rho_{\mathrm{D}^{N_0}}(x - x_m) + \sum_{n=1}^{N_0} b_n p_n(x) \tag{14}$$

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 09,2020 at 15:06:39 UTC from IEEE Xplore. Restrictions apply.

is a polynomial spline with knots at the data points $\{x_m\}_{m=1}^M$, and where $\{a_m\}_{m=1}^M$ and $\{b_n\}_{n=1}^{N_0}$ are suitable sets of coefficients. On setting $p = 2$, we recover the result given in [9]. Equations (13)–(14) show that the N_0th derivative of the solution to our continuous-domain problem lies in a finite-dimensional manifold. The solution itself can then be obtained by taking an N_0-fold integral, subject to adequate boundary conditions. However, for $p \neq 2$, we have a nonlinear mapping in (13). This makes it difficult to interpret other effects of regularization on the solution. Moreover, due to this nonlinear mapping, these solutions do not readily lend themselves to a discretization scheme, unlike in the gTV and L_2 cases. Therefore, we propose a spline-based discretization scheme to numerically solve the L_p-regularized generalized-interpolation problem for $p \geq 1$.

III. Polynomial Splines

Polynomial splines of degree N_0 form an essential component of our discretization scheme. They are piecewise-defined functions where each piece is a polynomial of degree N_0. These pieces are connected in a manner such that the first $(N_0 - 1)$ derivatives of the function are continuous. The points where the pieces are connected are called knots. A cardinal polynomial spline of degree N_0 has its knots on the integer grid and can be expressed uniquely in the form of a B-spline expansion [24]

$$f(x) = \sum_{k \in \mathbb{Z}} c[k] \beta_+^{N_0}(x - k), \tag{15}$$

where $\beta_+^{N_0}(x)$ is the causal B-spline of degree N_0 and $(c[k])_{k\in\mathbb{Z}}$ are the expansion coefficients. The causal B-spline of degree 0 is defined as:

$$\beta_+^0(x) = \begin{cases} 1, & \text{if } 0 \leq x < 1 \\ 0, & \text{otherwise,} \end{cases} \tag{16}$$

while the causal B-spline of degree N_0 is obtained by the $(N_0 + 1)$-fold convolution of $\beta_+^0(x)$ given by

$$\beta_+^{N_0}(x) = \underbrace{(\beta_+^0 * \beta_+^0 * \ldots * \beta_+^0)}_{N_0 \text{ convolutions}}(x). \tag{17}$$

We are interested in polynomial splines with knots located on a uniform grid of size h (in other words, the knots lie in $h\mathbb{Z}$). Such a spline of degree N_0 admits the B-spline expansion

$$f_h(x) = \sum_{k \in \mathbb{Z}} c_h[k] \beta_h^{N_0}(x - kh), \tag{18}$$

where $\beta_h^{N_0}(x) = \beta_+^{N_0}(\frac{x}{h})$ is the causal scaled B-spline of degree N_0. It is uniquely specified by its coefficients $c_h = (c_h[k])_{k\in\mathbb{Z}}$. We illustrate in Fig. 2 that $\beta_h^{N_0}(x)$ is compactly supported in $[0, (N_0 + 1)h]$. In fact, the B-spline $\beta_h^{N_0}(x)$ is the polynomial spline of degree N_0, with knots in $h\mathbb{Z}$, that has the shortest support [25].

Polynomial splines are closely linked to derivative operators of the form D^{N_0} $(N_0 \geq 1)$. The operator D^{N_0} is associated with the scaled B-spline of degree $(N_0 - 1)$ according to

$$\mathrm{D}^{N_0}\{\beta_h^{N_0-1}\}(x) = \frac{1}{h^{N_0-1}} \sum_{k \in \mathbb{Z}} d_{N_0}[k] \delta(x - kh). \tag{19}$$

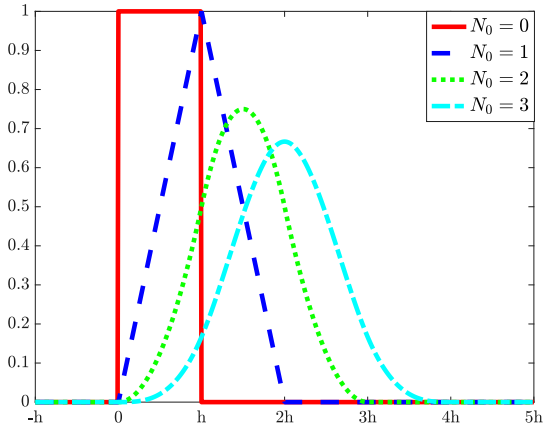

Fig. 2. Causal B-splines $\beta_h^{N_0}(x)$ with scaling factor h.

The sequence $(d_{N_0}[k])_{k\in\mathbb{Z}}$ is characterized by its z-transform

$$d_{N_0}(z) = (1 - z^{-1})^{N_0} \tag{20}$$

and is supported in $\{0, \ldots, N_0\}$. In Table I, we provide the explicit forms of $\beta_h^{N_0-1}(x)$ and $(d_{N_0}[k])_{k\in\mathbb{Z}}$ for $N_0 = 1, 2, 3$.

IV. Discretization Scheme

A. Search Space

We discretize the continuous-domain problem (12) by restricting the search space to a suitable space of polynomial splines, defined as

$$L_{p,h}^{N_0}(\mathbb{R}) = \left\{ \sum_{k \in \mathbb{Z}} c[k] \beta_h^{N_0}(\cdot - kh) \ : \ c \in \ell_p^{N_0}(\mathbb{Z}) \right\}, \tag{21}$$

where $\beta_h^{N_0}$ is the scaled B-spline of degree N_0, $h > 0$ is the grid size, and

$$\ell_p^{N_0}(\mathbb{Z}) = \{(c[k])_{k\in\mathbb{Z}} : (d_{N_0} * c) \in \ell_p(\mathbb{Z})\}. \tag{22}$$

The choice of the search space $L_{p,h}^{N_0}(\mathbb{R})$ is guided by its exact discretization property which we discuss in Section IV-B. Moreover, the approximation power of splines ensures that, when h is sufficiently small, the search space $L_{p,h}^{N_0}(\mathbb{R})$ contains functions that are arbitrarily close to the solution of the unrestricted continuous-domain problem (12). We present a detailed argument for this in Section IV-D. The fact that $L_{p,h}^{N_0}(\mathbb{R})$ is represented in a B-spline basis is another advantage. B-splines are compactly supported and form a Riesz basis [26], thus resulting in a well-conditioned discretization.

B. Exact Discretization

The exact discretization property of the function space $L_{p,h}^{N_0}(\mathbb{R})$ stems from Proposition 1.

Proposition 1: For any function $f \in L_{p,h}^{N_0}(\mathbb{R})$ with $p \in (0, \infty]$, we have that

$$\|\mathrm{D}^{N_0}\{f\}\|_{L_p} = \left\| \frac{1}{h^{N_0 - 1/p}} (d_{N_0} * c) \right\|_{\ell_p}. \tag{23}$$

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 09,2020 at 15:06:39 UTC from IEEE Xplore. Restrictions apply.

TABLE I
THE OPERATOR D^{N_0} AND THE SCALED B-SPLINE $\beta_h^{N_0-1}(x)$ AND SEQUENCE $(d_{N_0}[k])_{k\in\mathbb{Z}}$ ASSOCIATED WITH IT

$\mathrm{L}=\mathrm{D}^{N_0}$	$\beta_h^{N_0-1}(x)$	$(d_{N_0}[0],\ldots,d_{N_0}[N_0])$
D	$\beta_h^0(x)=\begin{cases}1, & 0\le x<h\\ 0, & \text{otherwise}\end{cases}$	$(1,-1)$
D^2	$\beta_h^1(x)=\begin{cases}x/h, & 0\le x<h\\ (2h-x)/h, & h\le x<2h\\ 0, & \text{otherwise}\end{cases}$	$(1,-2,1)$
D^3	$\beta_h^2(x)=\begin{cases}x^2/2h, & 0\le x<h\\ (-2x^2+6xh-3h^2)/2h^2, & h\le x<2h\\ (3h-x)^2)/2h^2, & 2h\le x<3h\\ 0, & \text{otherwise}\end{cases}$	$(1,-3,3,-1)$

Proof: A scaled B-spline of degree N_0 can be expressed as

$$\beta_h^{N_0}(x)=\frac{1}{h}(\beta_h^{N_0-1}*\beta_h^0)(x). \tag{24}$$

Using (19) and (24), we deduce that

$$\mathrm{D}^{N_0}\{\beta_h^{N_0}\}(x)=\frac{1}{h^{N_0}}\sum_{k\in\mathbb{Z}}d_{N_0}[k]\beta_h^0(x-kh). \tag{25}$$

Therefore, for any $f\in L_{p,h}^{N_0}(\mathbb{R})$ it stands that

$$\mathrm{D}^{N_0}\{f\}(x)=\frac{1}{h^{N_0}}\sum_{k\in\mathbb{Z}}(d_{N_0}*c)[k]\beta_h^0(x-kh). \tag{26}$$

Equation (26) implies that $\mathrm{D}^{N_0}\{f\}$ is a piecewise-constant function. For $p\in(0,\infty)$, the following holds:

$$\begin{aligned}\|\mathrm{D}^{N_0}\{f\}\|_{L_p}&=\left(\int_{\mathbb{R}}\Big|\frac{1}{h^{N_0}}\sum_{k\in\mathbb{Z}}(d_{N_0}*c)[k]\beta_h^0(x-kh)\Big|^p\,\mathrm{d}x\right)^{\frac{1}{p}}\\&=\left(\sum_{k\in\mathbb{Z}}h\Big|\frac{1}{h^{N_0}}(d_{N_0}*c)[k]\Big|^p\right)^{\frac{1}{p}}\\&=\left\|\frac{1}{h^{N_0-1/p}}(d_{N_0}*c)\right\|_{\ell_p}.\end{aligned} \tag{27}$$

For the case $p=\infty$, we have that

$$\begin{aligned}\|\mathrm{D}^{N_0}\{f\}\|_{L_\infty}&=\operatorname*{ess\,sup}_{x\in\mathbb{R}}\Big|\frac{1}{h^{N_0}}\sum_{k\in\mathbb{Z}}(d_{N_0}*c)[k]\beta_h^0(x-kh)\Big|\\&=\sup_{k\in\mathbb{Z}}\Big|\frac{1}{h^{N_0}}(d_{N_0}*c)[k]\Big|\\&=\left\|\frac{1}{h^{N_0}}(d_{N_0}*c)\right\|_{\ell_\infty}.\end{aligned} \tag{28}$$

■

On plugging the parametric form (21) of any function $f\in L_{p,h}^{N_0}(\mathbb{R})$ into Problem (12) and using Proposition 1, we obtain the equivalent discrete problem

$$\mathcal{S}_{p,h}=\operatorname*{arg\,min}_{c\in\ell_p^{N_0}(\mathbb{Z})}\left\|\frac{1}{h^{N_0-1/p}}(d_{N_0}*c)\right\|_{\ell_p}$$
$$\text{s.t.}\ \sum_{k\in\mathbb{Z}}c[k]\boldsymbol{\nu}(\beta_h^{N_0}(\cdot-kh))=\mathbf{y} \tag{29}$$

The important thing to note here is that Problem (29) is *exactly* equivalent to the continuous-domain problem (12) restricted to the search space $L_{p,h}^{N_0}(\mathbb{R})$. In other words, by solving the above discrete problem, we effectively find a solution to the restricted continuous-domain problem, which is given by $\sum_{k\in\mathbb{Z}}c^*[k]\beta_h^{N_0}(\cdot-kh)$ with $c^*\in\mathcal{S}_{p,h}$. As indicated by Proposition 1, this discretization scheme is also valid for L_p quasi-norm regularization with $p\in(0,1)$. However, these values of p correspond to non-convex problems.

Interestingly, the function space $L_{1,h}^{N_0}(\mathbb{R})$ can also be used for discretizing the gTV problem (11), which then also happens to be equivalent to the $p=1$ case.

Proposition 2: For any function $f\in L_{1,h}^{N_0}(\mathbb{R})$, we have that

$$\|\mathrm{D}^{N_0}\{f\}\|_{\mathcal{M}}=\|\mathrm{D}^{N_0}\{f\}\|_{L_1}. \tag{30}$$

Proof: Equation (26) implies that $\mathrm{D}^{N_0}\{f\}$ is piecewise-constant. Moreover, since $(d_{N_0}*c)\in\ell_1(\mathbb{Z})$, we conclude that $\mathrm{D}^{N_0}\{f\}\in L_1(\mathbb{R})$. The relationship between the $\mathcal{M}$-norm and L_1-norm (8) leads to (30). ■

By restricting the search space in (11) to $L_{1,h}^{N_0}(\mathbb{R})$ and using Propositions 1 and 2, we obtain the discrete problem (29) with $p=1$.

The salient and novel aspect of our method is the exact discretization of the continuous-domain problem. To the best of our knowledge, there is no prior work that discretizes L_p-regularized continuous-domain problems, with a general p, exactly. As mentioned earlier, the cases of $p=2$ and gTV have also been handled in [9], [23]. However, those discretization schemes have been specifically derived from representer theorems for L_2 and gTV regularization, and unlike the method proposed in this paper, are not applicable for other values of p.

C. Finite-Dimensional Problem

In practice, we assume that the measurement functionals ν_m are supported over a finite interval $I_T=[0,T]$. Consequently, a finite number of B-spline expansion coefficients are now involved in the constraint term in (29). We denote the set of the indices of these coefficients by $K=\{k_{\min},\ldots,k_{\max}\}$; the

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 09,2020 at 15:06:39 UTC from IEEE Xplore. Restrictions apply.

cardinality of this set is $|K| = N$. We now state Proposition 3, which has been adapted from Lemma 3 in [23].

Proposition 3: If the measurement functionals $\{\nu_m\}_{m=1}^M$ are supported in I_T, then a solution $c^* \in \mathcal{S}_{p,h}$ of Problem (29) is uniquely determined by the N coefficients $c^*|_K = (c^*[k_{\min}], \ldots, c^*[k_{\max}])$.

This result ensures that we only need to optimize over the N B-spline coefficients that affect the constraint (or data) term in (29). As described in [23], the expansion coefficients outside the interval of interest I_T can be set in a way such that all the regularization terms that they affect are nullified. This allows us to write the infinite-dimensional convolution in (29) as a matrix multiplication, leading to the finite-dimensional optimization problem

$$S_{p,h} = \underset{\mathbf{c}\in\mathbb{R}^N}{\arg\min} \|\mathbf{Lc}\|_{\ell_p} \text{ s.t. } \mathbf{Hc} = \mathbf{y}, \tag{31}$$

where the system matrix $\mathbf{H} : \mathbb{R}^N \to \mathbb{R}^M$ is

$$\mathbf{H} = \begin{bmatrix} \vdots & & \vdots \\ \boldsymbol{\nu}(\beta_h^{N_0}(\cdot - k_{\min}h)) & \cdots & \boldsymbol{\nu}(\beta_h^{N_0}(\cdot - k_{\max}h)) \\ \vdots & & \vdots \end{bmatrix}, \tag{32}$$

and the regularization matrix $\mathbf{L} : \mathbb{R}^N \to \mathbb{R}^{N-N_0}$ is

$$\mathbf{L} = \frac{1}{h^{N_0-\frac{1}{p}}} \begin{bmatrix} d_{N_0}[N_0] & \cdots & d_{N_0}[0] & 0 & \cdots & 0 \\ 0 & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 \\ 0 & \cdots & 0 & d_{N_0}[N_0] & \cdots & d_{N_0}[0] \end{bmatrix}. \tag{33}$$

The solutions $\mathbf{c}^* \in S_{p,h}$ and $c^* \in \mathcal{S}_{p,h}$ are related in the following manner: $\mathbf{c}^* = c^*|_K = (c^*[k_{\min}], \ldots, c^*[k_{\max}])$. Proposition 3 implies that the solution to Problem (29) can be uniquely determined from $\mathbf{c}^*$. Thus, we conclude that Problem (31) is equivalent to the continuous-domain problem (12) ((11), respectively) restricted to the search space $L_{p,h}^{N_0}(\mathbb{R})$ ($L_{1,h}^{N_0}(\mathbb{R})$, respectively), in the sense that the continuous-domain solution can be fully described by $\mathbf{c}^*$.

D. Effect of the Grid Size

So far, we have seen that the solutions to our continuous-domain problems, when restricted to $L_{p,h}^{N_0}(\mathbb{R})$, can be obtained by simply solving the finite problem (31). Now, we look at the influence of the grid size h on these solutions. We define a linear projection operator for the function space $L_{p,h}^{N_0}(\mathbb{R})$ as

$$\mathcal{P}_{L_{p,h}^{N_0}}\{f\}(x) = \sum_{k\in\mathbb{Z}} \left\langle f, \frac{1}{h}\tilde{\beta}^{N_0}\left(\frac{\cdot}{h} - k\right)\right\rangle \beta_+^{N_0}\left(\frac{x}{h} - k\right), \tag{34}$$

where $\tilde{\beta}^{N_0}$ is a (generalized) function such that

$$\left\langle \beta_+^{N_0}(\cdot - p), \tilde{\beta}^{N_0}(\cdot - q)\right\rangle = \delta(p - q). \tag{35}$$

The operator defined in (34) is a valid projection operator since it is idempotent. This can be shown by using the biorthonormality condition (35).

We now state Theorem 4, adapted from [27], which bounds the L_p-norm of the error between a function $f \in L_p^{(N_0)}(\mathbb{R})$ (the search space of the unrestricted continuous-domain problem, as defined in (9)) and its projection onto $L_{p,h}^{N_0}(\mathbb{R})$.

Theorem 4: Let $\mathcal{P}_{L_{p,h}^{N_0}}$ be a linear projection operator for $L_{p,h}^{N_0}(\mathbb{R})$, as defined in (34). When $p \in (1, \infty)$, the error of approximation for any $f \in L_p^{(N_0)}(\mathbb{R})$ is

$$\left\| f - \mathcal{P}_{L_{p,h}^{N_0}}\{f\}\right\|_{L_p} = \mathcal{O}(h^{N_0}). \tag{36}$$

For a small-enough grid size h, the error of approximation for any $f \in L_p^{(N_0)}(\mathbb{R})$ will be negligible. Therefore, our restricted search space $L_{p,h}^{N_0}(\mathbb{R})$ will contain functions (projections) which are arbitrarily close to the solution of the unrestricted continuous-domain problem. Finally, to compute the solution to the restricted continuous-domain problem, we only need to solve the finite problem (31).

V. Multiresolution Algorithm

In this section, we discuss a multiresolution algorithm that computes a solution with the desired precision by gradually making the grid finer.

A. Solving the Finite Problem for a Fixed Grid Size

We first discuss the algorithm that we use to solve finite-dimensional problems of the form (31). As constrained-optimization problems are typically harder to solve numerically compared to their unconstrained counterparts, to make the optimization easier we consider the unconstrained version of (31) given by

$$S'_{p,h} = \underset{\mathbf{c}\in\mathbb{R}^N}{\arg\min} \left(\|\mathbf{y} - \mathbf{Hc}\|_2^2 + \lambda\psi_p(\|\mathbf{Lc}\|_{\ell_p})\right) \tag{37}$$

where $\lambda \in \mathbb{R}^+$ is the regularization parameter and the function $\psi_p : \mathbb{R}^+ \to \mathbb{R}^+$ is defined as

$$\psi_p(x) = \begin{cases} x^p & \text{if } p \in [1, \infty), \\ x & \text{if } p = \infty, \end{cases} \tag{38}$$

Since ψ_p is monotonic over $\mathbb{R}^+$, the solution(s) to the constrained problem (31) can be obtained from (37) in the limit by taking $\lambda \to 0$. Thus, we propose to solve our finite-dimensional problem (31) by solving (37) with a very small value of λ.

The case $p = 2$ is special since then the optimization problem (37) is quadratic and can be solved directly without the need for an iterative algorithm. The unique solution in this scenario can be obtained by solving the linear system of equations $(\mathbf{H}^T\mathbf{H} + \lambda\mathbf{L}^T\mathbf{L})\mathbf{c}^* = \mathbf{H}^T\mathbf{y}$, which is obtained by setting to zero the gradient with respect to $\mathbf{c}$ of the cost functional in (37). This can be done by various methods, including direct matrix inversion.

For the values of $p \in [1, \infty] \setminus \{2\}$, we use the well-known ADMM [21] to solve Problem (37). The update rules for ADMM

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 09,2020 at 15:06:39 UTC from IEEE Xplore. Restrictions apply.

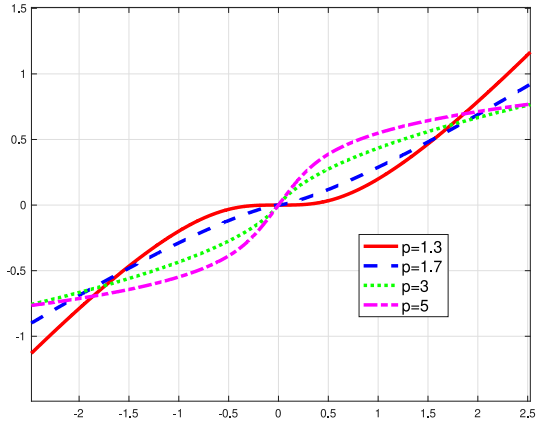

Fig. 3. Lookup tables for the proximal operators of $|\cdot|^p$.

in our case are

$$\mathbf{c}^{k+1} = \left(\mathbf{H}^T\mathbf{H} + \frac{\rho}{2}\mathbf{L}^T\mathbf{L}\right)^{-1}\left(\mathbf{H}^T\mathbf{y} + \frac{\rho}{2}\mathbf{L}^T(\mathbf{z}^k - \mathbf{u}^k)\right) \tag{39}$$

$$\mathbf{z}^{k+1} = \text{prox}_{\tilde{\lambda}\psi_p(\|\cdot\|_{\ell_p})}(\mathbf{L}\mathbf{c}^{k+1} + \mathbf{u}^k) \tag{40}$$

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \mathbf{L}\mathbf{c}^{k+1} - \mathbf{z}^{k+1}, \tag{41}$$

where $\mathbf{c}$ and $\mathbf{z}$ are the primal variables, $\mathbf{u}$ is the dual variable, $\rho > 0$ is the augmented-Lagrangian parameter and $\tilde{\lambda} = \lambda/\rho$. The proximal operator of a function g is defined as [28]

$$\text{prox}_g(\mathbf{x}) = \arg\min_{\mathbf{u}}\left(\frac{1}{2}\|\mathbf{u} - \mathbf{x}\|_2^2 + g(\mathbf{u})\right). \tag{42}$$

For $p = \{1, \infty\}$, the proximal operators involved in (40) have the closed-form expressions

$$\text{prox}_{\tilde{\lambda}\|\cdot\|_{\ell_1}}(\mathbf{x}) = \text{sgn}(\mathbf{x}) \otimes \max(|\mathbf{x}| - \tilde{\lambda}, 0) \tag{43}$$

$$\text{prox}_{\tilde{\lambda}\|\cdot\|_{\ell_\infty}}(\mathbf{x}) = \mathbf{x} - \tilde{\lambda}\text{proj}_{\|\cdot\|_{\ell_1}\leq 1}(\mathbf{x}/\tilde{\lambda}), \tag{44}$$

where the operators $\text{sgn}(\cdot)$ and $\max(\cdot)$ are applied component-wise, $\otimes$ denotes component-wise multiplication, and the projection operator is

$$\text{proj}_{\|\cdot\|_{\ell_1}\leq 1}(\mathbf{x}) = \arg\min_{\mathbf{u}:\|\mathbf{u}\|_{\ell_1}\leq 1}\|\mathbf{u} - \mathbf{x}\|_2^2. \tag{45}$$

This projector is computed as explained in [29]. Thus, the proximal operators can be computed efficiently for these two cases.

In general, we do not have a closed form expression for the proximal operator when $p \in (1, \infty)$. The additive separability of the function $\psi_p(\|\cdot\|_{\ell_p})$ can be used to observe that

$$[\text{prox}_{\tilde{\lambda}\psi_p(\|\cdot\|_{\ell_p})}(\mathbf{x})]_m = \text{prox}_{\tilde{\lambda}|\cdot|^p}([\mathbf{x}]_m). \tag{46}$$

Now, we only need to compute the proximal operator for the 1D function $\tilde{\lambda}|\cdot|^p : \mathbb{R} \to \mathbb{R}$, which we do with the help of lookup tables (LUTs). We provide in Fig. 3 a few examples of LUTs. An efficient implementation is achieved by exploiting properties of $\text{prox}_{\tilde{\lambda}|\cdot|^p}(\cdot)$ such as antisymmetry and monotonicity.

Algorithm 1: Multiresolution Algorithm.

Input: p, T, $\mathbf{y}$, $\boldsymbol{\nu}$, N_0, λ, h_{init}, ϵ.
Output: $\mathbf{c}^*$
Initialization: $\mathbf{c} = \mathbf{0}$, $t = 0$, rel_error $= \epsilon + 1$, prev_cost $= +\infty$
while rel_error $> \epsilon$ **do**
  $h = h_{\text{init}}/2^t$
  Update $\mathbf{H}$, $\mathbf{L}$
  if $p = 2$ **then**
    $\mathbf{c} = (\mathbf{H}^T\mathbf{H} + \lambda\mathbf{L}^T\mathbf{L})^{-1}\mathbf{H}^T\mathbf{y}$
  else
    $\mathbf{c} \leftarrow \text{ADMM}(\mathbf{c}_{\uparrow 2};\ p,\ \mathbf{y},\ \mathbf{H},\ \mathbf{L},\ \lambda)$
  rel_error $= |\text{cost}(\mathbf{c}) - \text{prev_cost}| \ /\ \text{prev_cost}$
  prev_cost $= \text{cost}(\mathbf{c})$
  $t \leftarrow t + 1$
if $p = 1$ **then**
  $\mathbf{y}_\lambda = \mathbf{H}\mathbf{c}$
  $\mathbf{c}^* = \text{Simplex}(\mathbf{y}_\lambda,\ \mathbf{H},\ \mathbf{L})$
else
  $\mathbf{c}^* = \mathbf{c}$

So far, we have seen that ADMM can be used to compute the unique solution to (37) when $p \in (1, \infty)$. When $p = \{1, \infty\}$, ADMM gives us one out of the possibly many solutions. In order to obtain a sparse solution for $p = 1$, we follow the procedure proposed in [9]. The solution $\mathbf{c}^* \in S'_{p,h}$ obtained via ADMM is used to generate the measurements $\mathbf{y}_\lambda = \mathbf{H}\mathbf{c}^*$. Using these "denoised" measurements, Problem (37) is then recast as a linear program which we solve using the simplex algorithm [30]. The simplex algorithm guarantees that we reach an extreme point of $S'_{1,h}$, which is sparse.

B. Grid Refinement

We begin with a coarse grid h_{init} and make it finer gradually until a further decrease of the grid size does not affect the solution much. At each iteration $t \in \mathbb{W}$, we pick a grid size $h_t = h_{\text{init}}/2^t$, splitting the grid from the previous iteration in half. We then solve the corresponding finite problem.

For this sequence of grid sizes, we observe that the search spaces are embedded like $L^{N_0}_{p,h_t}(\mathbb{R}) \subset L^{N_0}_{p,h_{t+1}}(\mathbb{R})$. This ensures that, by splitting the grid in half, we obtain a refined solution that is at least as good in terms of the cost function. Finally, we keep making the grid finer until the relative decrease in cost is less than some desired tolerance level ϵ. Another advantage of this embedding property is that the solution from the previous grid can be used as initialization for ADMM, which tends to improve the speed of convergence. This algorithm is adapted from the work in [23].

In Algorithm 1, $\mathbf{c}_{\uparrow 2}$ corresponds to the coefficients $\mathbf{c}$ modified to match a grid that is twice as fine as that of $\mathbf{c}$. The routine $\text{ADMM}(\mathbf{c}_{\uparrow 2};\ p,\ \mathbf{y},\ \mathbf{H},\ \mathbf{L},\ \lambda)$ runs ADMM on Problem (37) with $\mathbf{c}_{\uparrow 2}$ as the initialization while the routine $\text{Simplex}(\mathbf{y}_\lambda,\ \mathbf{H},\ \mathbf{L})$ runs the simplex algorithm on the linear program obtained from Problem (37) by using the denoised measurements $\mathbf{y}_\lambda$.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 09,2020 at 15:06:39 UTC from IEEE Xplore. Restrictions apply.

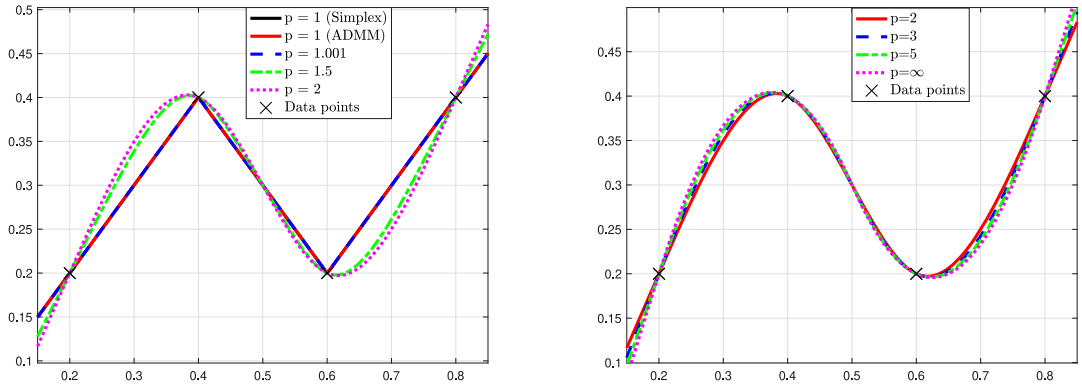

Fig. 4. Unique gTV solution ($\mathrm{L} = \mathrm{D}^2$). The simplex and ADMM solutions are coincident.

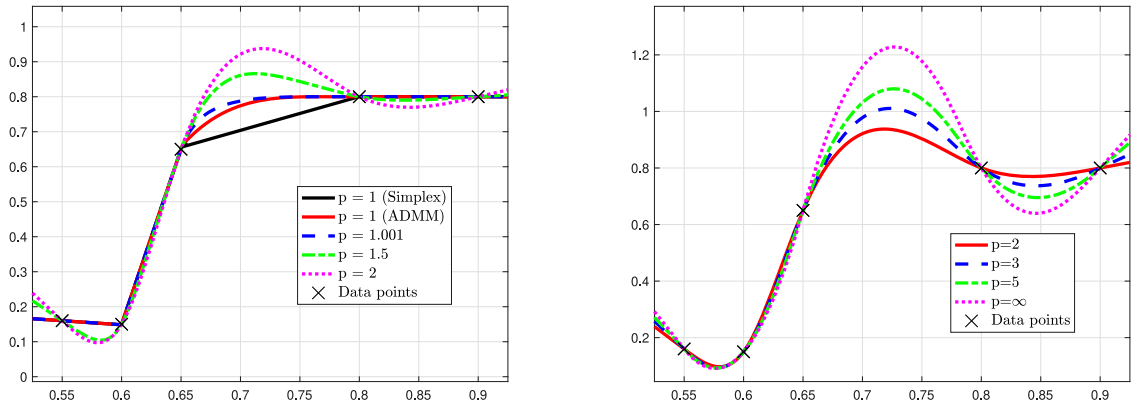

Fig. 5. Multiple gTV solutions ($\mathrm{L} = \mathrm{D}^2$).

VI. Numerical Experiments

We now present numerical results that allow us to identify certain properties of L_p-norm regularization and thus understand its effect. We have implemented our multiresolution algorithm using GlobalBioIm [31], a MATLAB library designed for solving inverse problems.

A. Setup

In our experiments, we have considered two types of measurement functionals.

- *Dirac Impulses:* In this setting, the given measurement operator takes the form $\boldsymbol{\nu}(f) = (\langle \delta(\cdot - x_1), f\rangle, \ldots, \langle \delta(\cdot - x_M), f\rangle) = (f(x_1), \ldots, f(x_M))$, where the points $\{x_m\}_{m=1}^M$ lie within the interval I_T. This operator corresponds to the standard interpolation problem that was discussed in Section I. We ensure that the points $\{x_m\}_{m=1}^M$ are pairwise distinct and that $M \geq N_0$, so that the operator $\boldsymbol{\nu}$ satisfies the condition $\mathcal{N}_{\mathrm{D}^{N_0}} \cap \mathcal{N}_{\boldsymbol{\nu}} = \{0\}$.
- *Dephased Cosines:* In this case, the measurement functionals are $\nu_1 = \mathbb{1}_{[0,T]}$ and $\nu_m = \cos(\omega_m x + \theta_m) \times \mathbb{1}_{[0,T]}$ for $m = \{2, 3, \ldots, M\}$. This operator corresponds to a variant of the Fourier interpolation problem which is relevant to magnetic resonance imaging. In order to construct such an operator and the corresponding measurements for our experiments, we first generated a function s and picked a threshold frequency $\omega_{\max}$ such that the spectrum of s had little energy above $\omega_{\max}$. The frequencies ω_m were then drawn uniformly at random from $(0, \omega_{\max}]$ while the phases θ_m were drawn uniformly at random from $[0, \pi)$. This operator $\boldsymbol{\nu}$ was applied to s to generate the measurements that we use in the experiments involving dephased cosines.

The regularization parameter was set to $\lambda = 10^{-10}$ in the first two experiments and $\lambda = 10^{-15}$ in the last two experiments. For all examples that we present in this section, the grid tolerance was set to $\epsilon = 10^{-3}$. In each example, we compute the solution for several values of $p \in [1, \infty]$.

B. Results

1) Continuum of Solutions & Sparsity: We first present two examples (Figs. 4 and 5) to talk about the behavior of the solution as the value of p is changed. In these examples, the measurement functionals are Dirac impulses (interpolation problem) and the regularization operator is $\mathrm{L} = \mathrm{D}^2$. Both examples show that, as we vary p from ∞ to 1 (note that $p = 1$ corresponds to the gTV case), the solutions gradually move towards the (or one of the) gTV solution(s). For the example in Fig. 4, the computed gTV solutions with and without applying the simplex are the same and resemble a linear spline with two knots, in agreement with [10]. It can be shown that this particular sparse solution is the

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 09,2020 at 15:06:39 UTC from IEEE Xplore. Restrictions apply.

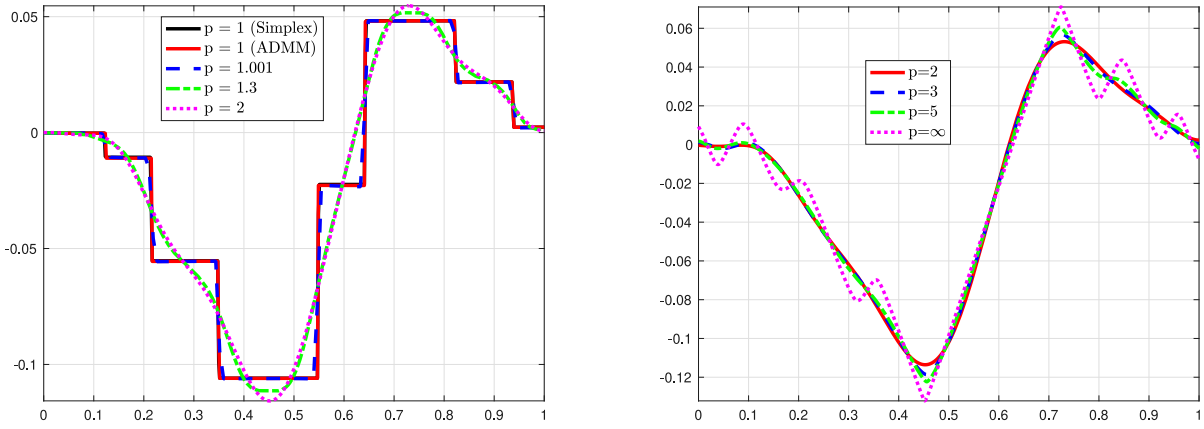

Fig. 6. Dephased-cosine measurement functionals ($\mathrm{L} = \mathrm{D}$, $M = 15$). For $p = 1$, the simplex and ADMM solutions are coincident.

unique solution to the gTV problem. In this case, we see that the solution for $p = 1.001$ is close to the unique sparse gTV solution.

By contrast, the configuration of the data points in Fig. 5 is such that the gTV problem has multiple solutions. This can be seen in the plots as the solution obtained by running the simplex after ADMM is sparse (linear spline with three knots), while the solution obtained via ADMM only is non-sparse. Interestingly in this case, the solution for $p = 1.001$ is close to a non-sparse gTV solution. Based on the above observations and additional experiments of the same nature, we make several claims.

- There exists a continuum of solutions when p is varied from ∞ to 1.
- When the gTV problem has a unique solution, the continuum converges to that unique sparse solution as $p \to 1$.
- When the gTV problem has multiple solutions, the continuum converges to one of its non-sparse solutions as $p \to 1$.

We discuss two implications of our claims. Firstly, the existence of a continuum implies that one can use L_p-norm regularization with $p \in (1, \infty)$, to "interpolate" between the properties of the gTV and L_∞ solutions. One such property is regularity or smoothness. In Figs. 4 and 5, we observe that the smoothness of the solution reduces as p decreases. Secondly, we conclude that L_p-norm regularization with a small p can be used as a sparsity-promoting prior in settings where the gTV solution is guaranteed to be unique. This is in line with the use of discrete ℓ_p-norm regularization, with a small p, in compressed-sensing frameworks.

As further illustration, we also provide an example with the dephased-cosine measurement functionals. In this case, the regularization operator was $\mathrm{L} = \mathrm{D}$, leading to a piecewise-constant gTV solution in Fig. 6. The continuum of solutions and change in regularity, as p is varied from ∞ to 1, is evident in this figure.

2) Gibbs-Like Oscillations: In the interpolation of step-like functions using splines, Gibbs-like oscillations are observed at the discontinuities [17], [32], [33]. We use the step and staircase functions (Fig. 7) to investigate this effect in our L_p-regularized problem. In these cases, we observe that the solutions exhibit an oscillatory behavior (with an overshoot at the discontinuity) which decreases as p goes from ∞ to 1. Moreover, as p becomes smaller, the oscillatory effect of the discontinuity becomes more localized. We claim that

- L_p-norm regularization with a smaller p results in weaker Gibbs-like oscillations at the edges.

We would like to point out that the above claims exclude the special case of spatial interpolation with $\mathrm{L} = \mathrm{D}$. Here, all values of $p \in (1, \infty)$ generate the same solution, which is a linear spline with knots at the data points. This can be inferred from the theoretical result stated in Section II.

3) Effect of N_0: We now discuss the influence of the operator $\mathrm{L} = \mathrm{D}^{N_0}$ which is the second component of our regularization term. In Fig. 8, we present an example where we fix $p = 1.5$ and compute the solutions for different values of N_0. Our general observation is that

- For any $p \in [1, \infty]$, the solution becomes smoother and exhibits more oscillations as N_0 increases.

4) Comparison With Shannon's Sinc Interpolation: Consider a standard interpolation problem with uniformly spaced points

$$x_m = m\Delta, \quad m = 1, 2, \ldots, M, \tag{47}$$

where $\Delta > 0$ is the spacing between any two consecutive points x_m, and measurements $\{y_m\}_{m=1}^{M}$. In this case, the well-known sinc interpolant is given by

$$f_{\text{sinc}}(x) = \sum_{m=1}^{M} y_m \, \text{sinc}\left(\frac{x - m\Delta}{\Delta}\right). \tag{48}$$

Remarkably, the variational formulation (12) of the above interpolation problem includes Shannon's sinc interpolation scheme as a special case corresponding to $p = 2$ and $N_0 \to \infty$ [34].

In many applications such as image scaling and image registration, smoother interpolating functions are desirable since they are well-behaved with well-defined multi-order derivatives. While $f_{\text{sinc}}(x)$ is a highly regular function, unfortunately it also exhibits strong Gibbs-like oscillations at sharp transitions. On the other hand, as observed in the previous experiments, by controlling the values of p and N_0, L_p-regularized solutions can be made to achieve a balance between smoothness and oscillatory behaviour.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 09,2020 at 15:06:39 UTC from IEEE Xplore. Restrictions apply.

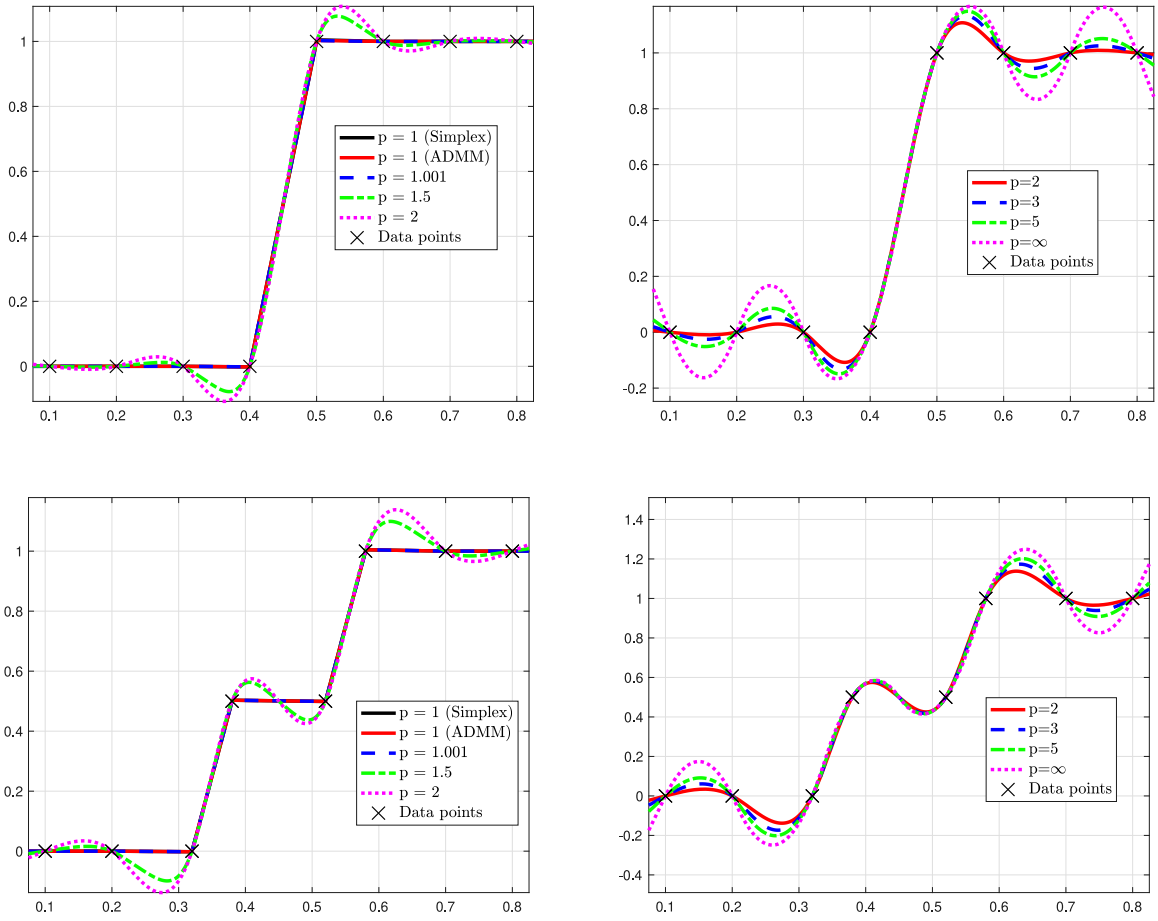

Fig. 7. Illustration of Gibbs-like oscillations ($\mathrm{L} = \mathrm{D}^2$). For $p = 1$, the simplex and ADMM solutions are coincident.

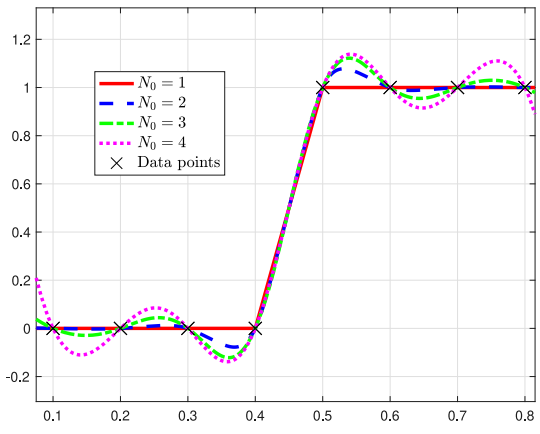

Fig. 8. Effect of the regularization operator D^{N_0} for a fixed $p = 1.5$.

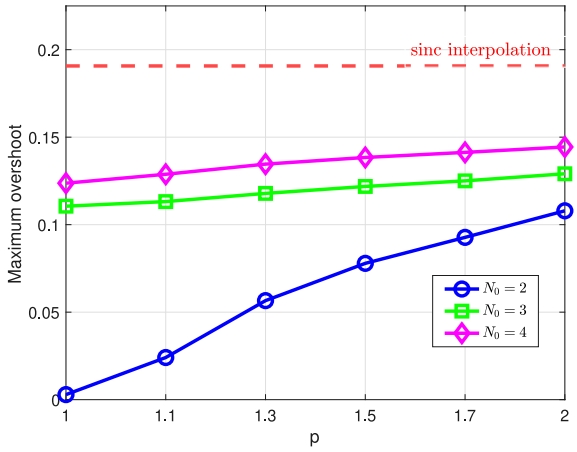

Fig. 9. Maximum overshoot values for interpolation of the data points from Fig. 8.

To illustrate this advantage of our framework, we consider interpolation of the data points from Fig. 8. We compute the maximum overshoot (which is related to the extent of the oscillations) of the sinc interpolant and the L_p-regularized interpolant for several values of p and N_0, and we plot the results in Fig. 9. For ease of comparison, we indicate the maximum overshoot for sinc interpolation, which is quite high, as a horizontal dashed line. The plots for the L_p-regularized solutions show that N_0 and p (more so when N_0 is small) can be varied to control the overshoots or oscillations, and balance them with the desired smoothness.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 09,2020 at 15:06:39 UTC from IEEE Xplore. Restrictions apply.

VII. Conclusion

We have implemented a multiresolution algorithm to solve numerically the generalized-interpolation problem with L_p-norm regularization, along with its unconstrained variants. We have shown that an appropriate grid-based B-spline basis can be used to exactly discretize the (restricted) continuous-domain problem. Based on previous results from approximation theory and splines, we have argued that as the grid size goes to zero, the computed solution approaches the solution of the unrestricted continuous-domain problem. With the help of numerical results in the context of spatial and Fourier interpolation, we have established the existence of a continuum of solutions as p goes from ∞ to 1. Finally, we have made insightful observations about properties of the L_p-regularized solutions such as sparsity, regularity, and Gibbs-like oscillations.

Appendix

Consider the unconstrained optimization problem in (2):

$$S = \arg\min_{f \in \mathcal{X}} \left(\underbrace{E\left(\mathbf{y}, \boldsymbol{\nu}(f)\right) + \lambda R(f)}_{J(f)} \right). \tag{49}$$

Here, we show that if E is strictly convex and R is convex, then all the solutions $f^* \in S$ generate the same measurement vector $\mathbf{z}_0 = \boldsymbol{\nu}(f^*)$. The proof is adapted from [35] and is based on standard arguments in convex analysis.

Let $f_1^*, f_2^* \in S$ be two solutions of (49) such that they produce different measurements i.e., $\boldsymbol{\nu}(f_1^*) \neq \boldsymbol{\nu}(f_2^*)$. Let the minimum value of the objective function be $J^* = J(f_1^*) = J(f_2^*)$. For a candidate function $f_c = \alpha f_1^* + (1-\alpha) f_2^*$, with $\alpha \in (0,1)$, we have

$$\begin{aligned} J(f_c) &= E\left(\mathbf{y}, \boldsymbol{\nu}\left(\alpha f_1^* + (1-\alpha) f_2^*\right)\right) + \lambda R\left(\alpha f_1^* + (1-\alpha) f_2^*\right) \\ &< \left(\alpha \left(\underbrace{E\left(\mathbf{y}, \boldsymbol{\nu}(f_1^*)\right) + \lambda R(f_1^*)}_{J^*} \right) \right. \\ &\quad \left. + (1-\alpha) \underbrace{E(\mathbf{y}, \boldsymbol{\nu}(f_2^*)) + \lambda R(f_2^*)}_{J^*} \right) = J^*. \end{aligned} \tag{50}$$

The above strict inequality is due to the fact that E is strictly convex and R is convex. The relation $J(f_c) < J^*$ is a contradiction and thus $\boldsymbol{\nu}(f_1^*) = \boldsymbol{\nu}(f_2^*) = \mathbf{z}_0$.

References

[1] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," *Soviet Math.*, vol. 4, pp. 1035–1038, 1963.

[2] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*. Boca Raton, FL, USA: CRC Press, 1998.

[3] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[4] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, vol. 1, no. 3. Basel, Switzerland: Birkhäuser Basel, 2013.

[5] M. Lustig, D. L. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Resonance Med.*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007.

[6] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

[7] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 707–710, Oct. 2007.

[8] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001.

[9] H. Gupta, J. Fageot, and M. Unser, "Continuous-domain solutions of linear inverse problems with Tikhonov *versus* generalized TV regularization," *IEEE Trans. Signal Process.*, vol. 66, no. 17, pp. 4670–4684, Sep. 2018.

[10] M. Unser, J. Fageot, and J. P. Ward, "Splines are universal solutions of linear inverse problems with generalized TV regularization," *SIAM Rev.*, vol. 59, no. 4, pp. 769–793, Dec. 2017.

[11] C. de Boor, "On "best" interpolation," *J. Approximation Theory*, vol. 16, no. 1, pp. 28–42, 1976.

[12] P. Copley and L. Schumaker, "On pLg-splines," *J. Approximation Theory*, vol. 23, no. 1, pp. 1–28, 1978.

[13] C. Micchelli, P. W. Smith, J. Swetits, and J. D. Ward, "Constrained L_p approximation," *Constructive Approximation*, vol. 1, no. 1, pp. 93–102, 1985.

[14] A. Pinkus, "On smoothest interpolants," *SIAM J. Math. Anal.*, vol. 19, no. 6, pp. 1431–1441, 1988.

[15] J. Favard, "On interpolation," *Bulletin de la Société Mathématique de France*, vol. 67, pp. 102–113, 1939.

[16] S. Karlin, "Interpolation properties of generalized perfect splines and the solutions of certain extremal problems. I," *Trans. Amer. Math. Soc.*, vol. 206, pp. 25–66, 1975.

[17] J. E. Lavery, "Univariate cubic L_p splines and shape-preserving, multiscale interpolation by univariate cubic L_1 splines," *Comput. Aided Geometric Des.*, vol. 17, no. 4, pp. 319–336, 2000.

[18] P. Auquiert, O. Gibaru, and E. Nyiri, "C^1 and C^2-continuous polynomial parametric L_p splines ($p \geq 1$)," *Comput. Aided Geometric Des.*, vol. 24, no. 7, pp. 373–394, 2007.

[19] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.: A J. Issued by the Courant Inst. Math. Sci.*, vol. 57, no. 11, pp. 1413–1457, 2004.

[20] I. Daubechies, M. Defrise, and C. De Mol, "Sparsity-enforcing regularisation and ISTA revisited," *Inverse Problems*, vol. 32, no. 10, 2016, Art. no. 104001.

[21] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[22] M. Unser and P. Tafti, *An Introduction to Sparse Stochastic Processes*. Cambridge, U.K.: Cambridge Univ. Press, 2014, p. 367.

[23] T. Debarre, J. Fageot, H. Gupta, and M. Unser, "B-spline-based exact discretization of continuous-domain inverse problems with generalized TV regularization," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4457–4470, Jul. 2019.

[24] I. Schoenberg, "Contributions to the problem of approximation of equidistant data by analytic functions: Part A.— On the problem of smoothing or graduation. A first class of analytic approximation formulae," *Quart. Appl. Math.*, vol. 4, no. 1, pp. 45–99, 1946.

[25] I. J. Schoenberg, *Cardinal Spline Interpolation*, vol. 12. Philadelphia, PA, USA: SIAM, 1973.

[26] M. Unser and T. Blu, "Cardinal exponential splines: Part I—Theory and filtering algorithms," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1425–1438, Apr. 2005.

[27] J. Lei, "L_p-approximation by certain projection operators," *J. Math. Anal. Appl.*, vol. 185, no. 1, pp. 1–14, 1994.

[28] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Berlin, Germany: Springer, 2011, pp. 185–212.

[29] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, "Efficient projections onto the l1-ball for learning in high dimensions," in *Proc. 25th Int. Conf. Mach. Learn.*, ser. ICML '08. New York, NY, USA: ACM, 2008, pp. 272–279.

[30] G. B. Dantzig, A. Orden, and P. Wolfe, "The generalized simplex method for minimizing a linear form under linear inequality restraints," *Pacific J. Math.*, vol. 5, no. 2, pp. 183–195, Oct. 1955.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 09,2020 at 15:06:39 UTC from IEEE Xplore. Restrictions apply.

[31] E. Soubies *et al.*, "Pocket guide to solve inverse problems with GlobalBioIm," *Inverse Problems*, vol. 35, no. 10, pp. 1–20, 2019.
[32] F. Richards, "A Gibbs phenomenon for spline functions," *J. Approximation Theory*, vol. 66, no. 3, pp. 334–351, 1991.
[33] Z. Zhang and C. F. Martin, "Convergence and Gibbs' phenomenon in cubic spline interpolation of discontinuous functions," *J. Comput. Appl. Math.*, vol. 87, no. 2, pp. 359–371, 1997.
[34] A. Aldroubi, M. Unser, and M. Eden, "Cardinal spline filters: Stability and convergence to the ideal sinc interpolator," *Signal Process.*, vol. 28, no. 2, pp. 127–138, Aug. 1992.
[35] R. J. Tibshirani, "The lasso problem and uniqueness," *Electron. J. Statist.*, vol. 7, pp. 1456–1490, 2013.

Pakshal Bohra received the B.Tech and M.Tech degrees in electrical engineering from the Indian Institute of Technology, Bombay, India, in 2018. He is currently working toward the Ph.D. degree with the Biomedical Imaging Group, Ecole Polytechnique Fédérale de Lausanne, Switzerland, under the direction of M. Unser. His research focuses on splines, stochastic processes and learning for inverse problems.

Michael Unser (Fellow, IEEE) received the M.S. (summa cum laude) and Ph.D. degrees in electrical engineering from Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in 1981 and 1984, respectively. He is a Professor and Director of EPFL's Biomedical Imaging Group, Lausanne, Switzerland. His primary area of investigation is biomedical image processing. He is internationally recognized for his research contributions to sampling theory, wavelets, the use of splines for image processing, stochastic processes, and computational bioimaging. He has published over 350 journal papers on those topics. He is the author with P. Tafti of the book *An introduction to sparse stochastic processes* Cambridge University Press 2014.

From 1985 to 1997, he was with the Biomedical Engineering and Instrumentation Program, National Institutes of Health, Bethesda USA, conducting research on bioimaging.

Dr. Unser has served on the editorial board of most of the primary journals in his field including the IEEE TRANSACTIONS ON MEDICAL IMAGING (associate Editor-in-Chief 2003-2005), IEEE TRANSACTIONS ON IMAGE PROCESSING, *Proceedings of IEEE, and SIAM Journal of Imaging Sciences*. He is the founding chair of the technical committee on Bio Imaging and Signal Processing (BISP) of the IEEE Signal Processing Society.

Prof. Unser is a fellow of the IEEE (1999), an EURASIP fellow (2009), and a member of the Swiss Academy of Engineering Sciences. He is the recipient of several international prizes including five IEEE-SPS Best Paper Awards, two Technical Achievement Awards from the IEEE (2008 SPS and EMBS 2010), the Technical Achievement Award from EURASIP (2018), and recently Career Achievement Award (IEEE EMBS 2020).

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 09,2020 at 15:06:39 UTC from IEEE Xplore. Restrictions apply.