B-Spline-Based Exact Discretization of Continuous-Domain Inverse Problems with Generalized TV Regularization

Thomas Debarre, Julien Fageot, Harshit Gupta, and Michael Unser, *Fellow, IEEE*

***Abstract*—We study continuous-domain linear inverse problems with generalized Total-Variation (gTV) regularization, expressed in terms of a regularization operator L. It has recently been proved that such inverse problems have sparse spline solutions, with fewer jumps than the number of measurements. Moreover, the type of spline solely depends on L (L-splines), and is independent of the measurements. The continuous-domain inverse problem can be recast in an exact way as a finite-dimensional problem by restricting the search space to splines with knots on a uniform finite grid. However, expressing the L-spline coefficients in the dictionary basis of the Green's function of L is ill-suited for practical problems due to its infinite support. Instead, we propose to formulate the problem in the B-spline dictionary basis, which leads to better-conditioned problems. As we make the grid finer, we show that a solution of the continuous-domain problem can be approached arbitrarily closely with functions of this search space. This result motivates our proposed multiresolution algorithm, which computes sparse solutions of our inverse problem. We demonstrate that this algorithm is computationally feasible for 1D signals when L is an ordinary differential operator.**

***Index Terms*—inverse problems, total variation, sparsity, compressed sensing, B-splines**

I. Introduction

The task in an inverse problem is to recover an unknown signal s from its (usually noise-corrupted) measurements $\mathbf{y}$, which are acquired following a forward model $\mathbf{y} \approx \boldsymbol{\nu}(s)$, *e.g.*, Fourier samples in the case of MRI data. In many cases, such problems are ill-posed in the sense that many different signals yield identical measurements. The ill-posedness of inverse problems can be circumvented by using a regularization term, the choice of which is guided by our prior knowledge of the underlying signal.

In recent years, ℓ_1 regularization has become increasingly popular with the surge of compressed sensing [1]–[3] for the reconstruction of discrete signals. Its benefits have been extensively documented, including its sparsity-promoting effect [4], [5], its perfect recovery properties under certain conditions [1], [6], [7], the availability of efficient solvers [8], [9] or its apparent superiority over Tikhonov ℓ_2 regularization to recover many real-world signals [10].

Since many real-world signals are continuously defined, some recent papers are directed towards solving continuous-domain inverse problems, including [11]–[14]. However, for obvious considerations of computational feasibility, the vast majority of research efforts in the field of compressed sensing focus on discrete inverse problems. The standard approach is to express a continuous-domain signal s in a certain finite basis $\{\phi_n\}_{1\leq n\leq N}$ (typically a pixel basis), *i.e.*, $s = \sum_{n=1}^{N} c_n \phi_n$. Given M measurements $\mathbf{y} \in \mathbb{R}^M$, the inverse problem is then formulated in terms of the coefficients $\mathbf{c} = (c_1, \ldots, c_N)$ in a penalized form as

$$\min_{\mathbf{c}\in\mathbb{R}^N} \|\mathbf{H}\mathbf{c} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{L}\mathbf{c}\|_1 \tag{1}$$

where $\mathbf{H} : \mathbb{R}^N \to \mathbb{R}^M$ is the system matrix (discrete forward model), $\mathbf{L} : \mathbb{R}^N \to \mathbb{R}^N$ is the regularization matrix and λ is a regularization parameter. The choice of $\mathbf{L}$ allows us to promote the sparsity of $\mathbf{c}$ in a chosen transform domain, *e.g.*, a finite difference matrix (discrete TV regularization) or wavelets. However, there are several downsides to this discrete approach: the choice of the basis functions ϕ_n is guided by computational considerations, and is not necessarily matched to the characteristics of the underlying continuous-domain signal. Moreover, the discrete forward model $\mathbf{H}$ is often an approximation of its continuous counterpart (*e.g.*, the discrete Fourier transform for the continuous Fourier transform), which introduces discretization errors.

A. Continuous-Domain Framework

To address these limitations, we focus directly on 1D continuous-domain (*i.e.*, $s : \mathbb{R} \to \mathbb{R}$) inverse problems with gTV regularization using the framework of [15]:

$$\min_{s} \|\boldsymbol{\nu}(s) - \mathbf{y}\|_2^2 + \lambda \|\mathrm{L}\{s\}\|_{\mathcal{M}} \tag{2}$$

where $\boldsymbol{\nu}$ is a linear measurement operator (continuous forward model) and $\mathbf{y} \in \mathbb{R}^M$ are the noise-corrupted measurements. The regularization is defined in terms of a suitable operator L with Green's function ρ_{L}; the $\|\cdot\|_{\mathcal{M}}$ norm is a generalization of the L_1 norm, and is the continuous counterpart of the ℓ_1 norm. The prototypical example is the derivative operator $\mathrm{L} = \mathrm{D}$, leading to TV regularization. This paper relies on the main result of [15], a representer theorem (Theorem 1) which stems

This research was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, Grant 692726-GlobalBioIm, and by the Swiss National Science Foundation, Grant P2ELP2_181759.

The authors are with the Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (e-mail: thomas.debarre@epfl.ch; julien.fageot@epfl.ch; harshit.gupta@epfl.ch; michael.unser@epfl.ch). J. Fageot is also affiliated with Harvard University, Cambridge, MA 02138, USA

from the pioneer work of Fisher and Jerome [16]. This theorem states that Problem (2) has L-spline solutions of the form

$$s(x) = \sum_{k=1}^{K} a_k \rho_{\mathrm{L}}(x - x_k) + \sum_{n=1}^{N_0} b_n p_n(x), \tag{3}$$

where $K \leq M - N_0$, $a_k, x_k \in \mathbb{R}$ and $\{p_n\}_{n=1}^{N_0}$ form a basis of the null space of L. This result resonates with the sparsity-promoting effect of the ℓ_1 norm for discrete problems, since these spline solutions are sparse in the Green's function dictionary basis $\{\rho_{\mathrm{L}}(\cdot - \tau)\}_{\tau \in \mathbb{R}}$. For example, for $\mathrm{L} = \mathrm{D}$, a Green's function is the Heaviside step function $\rho_{\mathrm{L}} = {}_+$, which implies that s is a sparse piecewise-constant signal.

B. Green's Function Discretization

The form of the solutions (3) provides a natural basis to discretize Problem (2). As demonstrated in [17], by using basis functions $\{p_n\}_{n=1}^{N_0}$ and $\{\rho_{\mathrm{L}}(\cdot - x_n)\}_{n=1}^{N}$ where the knot x_n lie on a uniform finite grid, we get the following discrete optimization problem:

$$\min_{(\mathbf{a},\mathbf{b}) \in \mathbb{R}^{N+N_0}} \|\mathbf{H}_{\rho_{\mathrm{L}}}\mathbf{a} + \mathbf{H}_p\mathbf{b} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{a}\|_1 \tag{4}$$

with system matrices $\mathbf{H}_{\rho_{\mathrm{L}}} : \mathbb{R}^N \to \mathbb{R}^M$ and $\mathbf{H}_p : \mathbb{R}^{N_0} \to \mathbb{R}^M$. This problem is of the form (1), and can thus be solved using off-the-shelf convex optimization algorithms. The major asset of this approach is that the discrete problem is *exactly* equivalent to the underlying continuous problem restricted to the search space spanned by the basis functions. By making the grid finer, this search space includes functions arbitrarily close to solutions (3) of the full continuous-domain problem. However, the Green's function usually has infinite support (*e.g.* $\rho_{\mathrm{D}} = \mathbb{1}_+$), which makes the Green's function basis ill-suited for practical problems. In particular, Problem (4) is severely ill-conditioned, making the convergence of solvers slow and potentially numerically unstable.

C. Our Approach: B-Spline Discretization

We therefore propose to improve this discretization method by using an equivalent dictionary basis consisting of shifted B-splines, *i.e.*, $\{\beta_{\mathrm{L}}(\cdot - x_n)\}_{n=1}^{N_0}$ where β_{L} is the B-spline of L, and the knots x_n lie on a uniform finite grid. B-splines are popular signal processing tools [18]–[20], notably due to their finite support (*e.g.*, $\beta_{\mathrm{D}} = \mathbb{1}_{[0,1]}$). This basis leads to the following discrete optimization problem:

$$\min_{\mathbf{c} \in \mathbb{R}^N} \|\mathbf{H}_{\beta_{\mathrm{L}}}\mathbf{c} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{L}\mathbf{c}\|_1 \tag{5}$$

where $\mathbf{L}$ is a finite difference-like regularization matrix. This problem is of the same form as standard discrete compressed sensing-type problems (1), with the advantage that the chosen basis is matched to the form of the continuous-domain solution (3). Moreover, Problem (5) shares the exact discretization property of the Green's function basis, since (setting aside boundary issues) both bases are equivalent. However, the finite support of the B-splines makes this basis better suited for practical applications, and it induces well-conditioned problems. This leads to a rapid convergence of solvers for Problem (5), a prediction which will be confirmed by our experimental results.

D. Related Works

The key feature of our approach is that contrary to standard formulations, our discretization is *exact* in the continuous domain. Other approaches stemming from [21] have been undertaken in the litterature to solve continuous-domain inverse problems involving the TV norm (or related norms, *e.g.*, atomic norms in [22]). For instance, Adcock and Hansen have introduced the theory of infinite dimensional compressed sensing [13], [23]. Most of the research effort in this domain has been dedicated to formulating inverse problems in spaces of measures (typically to recover sums of Dirac impulses) and devising grid-free numerical algorithms to solve them, *e.g.*, [11], [12], [14], [24]–[30]. These grid-free approaches aim to recover the exact locations x_k of the jumps at super-resolution, using sophisticated algorithms based on duality. Although this is a sensible objective when the reconstructed signals consists of Dirac impulses, in our spline-based framework, finding the exact locations of the jumps is less critical since the reconstructed signals are smoother. We therefore take the stance of using a grid and B-splines which leads to a simple and effective algorithm, at the expense of (arbitrarily small) localization errors on the jumps.

E. Outline and Contributions

In this paper, we focus on the widely-used class of ordinary differential regularization operators L, which lead to exponential B-splines [31]. To set the scene, we present some background information on the continuous-domain inverse problem framework of [15] (Section II) and exponential B-splines (Section III). In terms of contribution, this paper extends [17] and parts of our experimental pipeline are adapted from this work. However, the use of B-splines is a critical improvement which leads to the following contributions:

- In Section IV, we define the search space consisting of L-splines with knots on a uniform grid and show that it has an equivalent formulation in the B-spline basis;
- In Section V, we show that the continuous inverse problem can be recast as a finite-dimensional problem of the form (1) in an *exact* way. We also demonstrate that the algorithm introduced in [17] can be adapted to our framework, and that it yields sparse solutions (with lower sparsity than in [17]);
- In Section VI, we prove that the optimal cost of the discrete problem converges to that of the continuous problem when the grid size goes to zero. We then use this result to devise a multiresolution algorithm which refines the grid until the desired level of accuracy is met (termination criterion);
- In Section VII, we demonstrate experimentally the effectiveness of our algorithm using different measurement types (ideal sampling in the spatial and Fourier domains). We also show that it compares favorably with standard purely discrete methods.

II. Continuous-Domain Inverse Problem

In this section, we present a class of continuous-domain inverse problems with gTV regularization. The solution sets

of such problems are described by a representer theorem introduced in [15], which is the backbone of this paper. We provide our reader the minimum knowledge required to understand the framework of [15]. For more information, we refer to the original article.

To introduce notations, we summarize the aim of our inverse problem, which is to recover a certain continuous-domain signal $s : \mathbb{R} \to \mathbb{R}$ given M noisy measurements $\mathbf{y} = \boldsymbol{\nu}(s) + \mathbf{n} \in \mathbb{R}^M$. The noiseless measurements $\boldsymbol{\nu}(s)$ are acquired through M linear measurement functionals $\boldsymbol{\nu} = (\nu_1, \dots, \nu_M)$, *i.e.*, $\boldsymbol{\nu}(s) = (\langle \nu_1, s \rangle, \dots, \langle \nu_M, s \rangle)$ ($\langle \nu_m, s \rangle$ stands for the duality product, which is given by $\int_{\mathbb{R}} \nu_m(x) s(x) dx$ when ν_m and s are ordinary functions). The ν_m functionals constitute the (known) forward model. The measurements are assumed to be corrupted by some additive noise $\mathbf{n}$.

A. Definitions

Let $\mathcal{S}'(\mathbb{R})$ be the space of tempered distributions, defined as the dual of the Schwartz space $\mathcal{S}(\mathbb{R})$ of smooth and rapidly decaying functions on $\mathbb{R}$. A typical example of a tempered distribution is the Dirac distribution δ, which is defined in terms of its dual product with a test function $\phi \in \mathcal{S}(\mathbb{R})$ by $\langle \delta, \phi \rangle = \phi(0)$.

The key element of our formulation is the regularization operator L. Throughout this paper, we will focus on continuous linear shift-invariant (LSI) operators $\mathrm{L} : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$. Such an operator is equivalent to a convolution and is conveniently represented by its frequency response $\widehat{L}(\omega)$. In fact, a full characterization in terms of frequency response is given by Schwarz in [32, Chapter 7, §5]. In short, for any $f \in \mathcal{S}'(\mathbb{R})$, we can write $\mathcal{F}\{\mathrm{L}\{f\}\}(\omega) = \widehat{L}(\omega)\widehat{f}(\omega)$ where $\omega \mapsto \widehat{L}(\omega)$ and its successive derivatives are smooth functions of slow growth (*i.e.*, bounded by a polynomial). Here, $\mathcal{F}$ denotes the generalized Fourier transform and $\widehat{f} = \mathcal{F}\{f\}$.

In order to be acceptable in our framework, L also needs to be *spline-admissible* in the following sense.

Definition 1 (Spline-admissible operator)**.** *A continuous LSI operator* $\mathrm{L} : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$ *is* spline-admissible *if it verifies the following properties*

- *there exists a locally integrable function of slow growth* $\rho_{\mathrm{L}} : \mathbb{R} \to \mathbb{R}$ *(the Green's function of* L*) which satisfies* $\mathrm{L}\{\rho_{\mathrm{L}}\} = \delta$*;*
- *its null space* $\mathcal{N}_{\mathrm{L}} = \{f \in \mathcal{S}'(\mathbb{R}) : \mathrm{L}\{f\} = 0\}$ *has finite dimension* N_0*.*

Note that Definition 1 is less general than that of [15] and Chapter 5 in [33], which include for example fractional operators or multi-dimensional operators such as the Laplacian. The present restriction is justifiable by the fact that the latter category of operators leads to B-splines with non-compact support, which are less convenient for our purpose.

The prototypical example of a spline-admissible operator is the derivative $\mathrm{L} = \mathrm{D}$, whose frequency response is $\widehat{L}(\omega) = \mathrm{j}\omega$ and whose causal Green's function is $\rho_{\mathrm{L}} = \mathbb{1}_+$. However, the Green's function is non unique, since adding any element of the null space $\mathcal{N}_{\mathrm{L}}$ (constant functions in this case) to a Green's function yields other valid Green's functions.

Definition 2 (Non-uniform L-spline)**.** *Let* L *be a spline-admissible operator in the sense of Definition 1. A* non-uniform L-spline *is a function* $s : \mathbb{R} \mapsto \mathbb{R}$ *verifying*

$$\mathrm{L}\{s\}(x) = \sum_{k \in \mathbb{Z}} a[k] \delta(x - x_k) \tag{6}$$

where $a[k] \in \mathbb{R}$ *is the amplitude of the* k*-th jump, and the* $x_k \in \mathbb{R}$ *are the pairwise distinct knot locations. The distribution* $w = \sum_{k \in \mathbb{Z}} a[k] \delta(\cdot - x_k)$ *is known as the* innovation *of the spline.*

It follows from Definition 2 that a non-uniform spline can equivalently be defined as

$$s(x) = p(x) + \sum_{k \in \mathbb{Z}} a[k] \rho_{\mathrm{L}}(x - x_k) \tag{7}$$

where $p \in \mathcal{N}_{\mathrm{L}}$. The ground truth signal in Figure 1 is an example of a non-uniform spline for $\mathrm{L} = \mathrm{D}$ (piecewise-constant functions). The locations of the jumps correspond to the x_k, and their amplitudes are the $a[k]$ coefficients.

Next, let $\mathcal{M}(\mathbb{R}) \subset \mathcal{S}'(\mathbb{R})$ be the space of finite Radon measures, which is known by the Riesz-Markov theorem [34, Chapter 6] to be the continuous dual of $C_0(\mathbb{R})$. The latter is the space of continuous functions vanishing at infinity, which is a Banach space when equipped with the supremum norm $\|\cdot\|_\infty$. Its dual space $\mathcal{M}(\mathbb{R})$ is therefore a Banach space equipped with the corresponding dual norm $\|\cdot\|_{\mathcal{M}}$, which is defined for $w \in \mathcal{M}(\mathbb{R})$ as

$$\|w\|_{\mathcal{M}} = \sup_{\phi \in C_0(\mathbb{R}), \|\phi\|_\infty = 1} \langle w, \phi \rangle. \tag{8}$$

The $\|\cdot\|_{\mathcal{M}}$ norm can be seen as a generalization of the L_1 norm: for any $f \in L_1(\mathbb{R})$, we have $\|f\|_{L_1} = \|f\|_{\mathcal{M}}$. Another crucial property for our application is the inclusion of shifted Dirac impulses in $\mathcal{M}(\mathbb{R})$, with $\|\delta(\cdot - x_0)\|_{\mathcal{M}} = 1$.

Since the $\mathcal{M}$ norm is used for regularization purposes in combination with L in our setting, we only consider functions f such that $\|\mathrm{L}\{f\}\|_{\mathcal{M}}$ is well defined. In all that follows, we will therefore consider the following native space of L

$$\mathcal{M}_{\mathrm{L}}(\mathbb{R}) = \Big\{ f \in \mathcal{S}'(\mathbb{R}) : \mathrm{L}\{f\} \in \mathcal{M}(\mathbb{R}) \Big\}, \tag{9}$$

which can be endowed with a Banach direct-sum topology [15, Theorem 5].

A crucial observation is that non-uniform L-splines as in (7) are included in $\mathcal{M}_{\mathrm{L}}(\mathbb{R})$ when $a \in \ell_1(\mathbb{Z})$, since $\|\mathrm{L}\{s\}\|_{\mathcal{M}} = \|\sum_{k \in \mathbb{Z}} a[k] \delta(\cdot - x_k)\|_{\mathcal{M}} = \|a\|_1$.

B. Representer Theorem

Now that all the relevant concepts have been introduced, we can state the representer theorem of [15] in an equivalent form formulated in [17] and in the case of a quadratic data fidelity cost function:

Theorem 1 (Continuous-Domain Representer Theorem)**.** *Let* L *be a spline-admissible operator in the sense of Definition 1, and let* $\boldsymbol{\nu} = (\nu_1, \dots, \nu_M) : \mathcal{M}_{\mathrm{L}}(\mathbb{R}) \to \mathbb{R}^M$ *be a weak*-continuous linear measurement operator composed of* $M \geq N_0$ *linear functionals* $\nu_m : f \mapsto \nu_m(f) \in \mathbb{R}$*. Assume that the*

intersection of the null spaces of L *and* $\boldsymbol{\nu}$ *is restricted to* $\{0\}$, i.e., $\mathcal{N}_{\boldsymbol{\nu}} \cap \mathcal{N}_{\mathrm{L}} = \{0\}$ *(well-posedness assumption). Then the linear inverse problem*

$$\mathcal{S} = \underset{f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R})}{\arg\min} \underbrace{\|\boldsymbol{\nu}(f) - \mathbf{y}\|_2^2 + \lambda \|\mathrm{L}\{f\}\|_{\mathcal{M}}}_{\mathcal{J}(f)} \tag{10}$$

has a non-empty weak-compact convex solution set* $\mathcal{S}$ *whose extreme points are non-uniform* L*-splines (in the sense of Definition 2) of the form*

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

where $\{p_n\}_{n=1,\dots,N_0}$ *form a basis of* $\mathcal{N}_{\mathrm{L}}$*,* $a_k, x_k \in \mathbb{R}$*, and the sparsity index* K *verifies* $K \leq M - N_0$*.*

The remarkable outcome of this theorem is that solutions of a continuous-domain problem with infinitely many degrees of freedom can be expressed with a small number $K + N_0 \leq M$ of coefficients. Moreover, the underlying infinite-dimensional dictionary basis $\{\rho_{\mathrm{L}}(\cdot - \tau)\}_{\tau \in \mathbb{R}}$ is fully determined by L and is thus independent of the measurements. This result confirms the sparsity-promoting effect of the L_1 norm (as a particular case of the $\|\cdot\|_{\mathcal{M}}$ norm), which is well-known in discrete problems, but much less so in continuous ones. It also confirms the usefulness of splines as bridges between the continuous and discrete worlds.

The parametric form of the solutions given by Theorem 1 makes it tempting to recast the continuous-domain problem into a discrete problem by feeding the parametric solution of (11) into the optimization Problem (10) and optimizing over the parameters a_k, x_k and b_k. Although this optimization problem is non-convex with respect to the knot locations x_k, this issue can be avoided by gridding. This amounts to restricting the search space of the problem to the space L-splines with knots on a uniform grid, which we properly define in Section IV.

III. Exponential Splines

For the sake of clarity, we restrict ourselves in all that follows to the class of ordinary differential regularization operators, which cover the vast majority of 1D real-world applications. These lead to so-called exponential splines, which have been studied extensively in [31]. In this section, we introduce basic information on exponential splines.

A. Differential Operators

We focus on ordinary differential operators

$$\mathrm{L} = \mathrm{D}^{N_0} + a_{N_0-1}\mathrm{D}^{N_0-1} + \ldots + a_0\mathrm{I} \tag{12}$$

where D and I are the derivative and identity operators respectively. Let

$$P(X) = X^{N_0} + \sum_{n=0}^{N_0-1} a_n X^n = \prod_{n=1}^{N_0} (X - \alpha_n)$$

be the polynomial function associated to L which has roots $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{N_0}) \in \mathbb{C}^{N_0}$. We can thus refer to L as $\mathrm{L}_{\boldsymbol{\alpha}}$. Let $\alpha_{(1)}, \ldots, \alpha_{(N_d)}$ be the distinct roots of P with multiplicity $n_{(1)}, \ldots, n_{(N_d)}$ ($N_d \leq N_0$). It is well known from linear system theory that the null space of $\mathrm{L}_{\boldsymbol{\alpha}}$ is

$$\mathcal{N}_{\boldsymbol{\alpha}} = \mathrm{span}\Big\{x \mapsto x^n \mathrm{e}^{\alpha_{(m)}x}\Big\}_{1 \leq m \leq N_d,\ 0 \leq n \leq n_{(m)}-1}. \tag{13}$$

For convenience, we characterize these operators in the Fourier domain as

$$\widehat{L}_{\boldsymbol{\alpha}}(w) = \prod_{n=1}^{N_0} (\mathrm{j}w - \alpha_n). \tag{14}$$

Among the possible Green's functions of $\mathrm{L}_{\boldsymbol{\alpha}}$, we select the canonical solution

$$\rho_{\boldsymbol{\alpha}}(x) = \mathcal{F}^{-1}\left\{\frac{1}{\widehat{L}_{\boldsymbol{\alpha}}(w)}\right\}(x). \tag{15}$$

which is uniquely defined. For example, the Green's function of the elementary operator $\mathrm{L}_\alpha = \mathrm{D} - \alpha I$ when $\mathrm{Re}(\alpha) < 0$ is

$$\rho_\alpha(x) = \mathbb{1}_+(x)\mathrm{e}^{\alpha x}. \tag{16}$$

B. Exponential B-splines

The exponential B-spline with knot spacing $h > 0$ of the differential operator $\mathrm{L}_{\boldsymbol{\alpha}}$ is defined as

$$\beta_{\boldsymbol{\alpha},h}(x) = \frac{1}{h^{N_0-1}} \mathcal{F}^{-1}\left\{\prod_{n=1}^{N_0} \frac{1 - \mathrm{e}^{h(\alpha_n - \mathrm{j}\omega)}}{\mathrm{j}\omega - \alpha_n}\right\}(x). \tag{17}$$

For example, the first-order exponential B-spline is

$$\beta_{\alpha,h}(x) = \mathbb{1}_{[0,h]}(x)\mathrm{e}^{\alpha x}, \tag{18}$$

and is thus supported in $[0, h]$. For higher orders, $\beta_{\boldsymbol{\alpha},h}$ is supported in $[0, N_0 h]$ since it is proportional to the convolution of N_0 first-order exponential B-splines. A simple Fourier calculation yields the innovation of the exponential B-spline

$$\mathrm{L}\{\beta_{\boldsymbol{\alpha},h}\}(x) = \frac{1}{h^{N_0-1}} \sum_{k=0}^{N_0} d_{h\boldsymbol{\alpha}}[k]\delta(x - hk) \tag{19}$$

where the sequence $d_{h\boldsymbol{\alpha}}$ is characterized by its z transform

$$D_{h\boldsymbol{\alpha}}(z) = \prod_{n=1}^{N_0} (1 - \mathrm{e}^{h\alpha_n} z^{-1}), \tag{20}$$

and thus has a finite support $\{0, \ldots, N_0\}$.

IV. Specification of the Search Space

A. Green's Function Representation

As explained earlier, in order to discretize the continuous-domain problem, we restrict the search space to L-splines (7) with knots on a uniform grid, *i.e.*, $x_k \in h\mathbb{Z}$ where $h > 0$ is the step size of the grid. In the case of exponential splines (*i.e.*, $\mathrm{L} = \mathrm{L}_{\boldsymbol{\alpha}}$), the search space with step size h is thus defined as

$$\mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R}) = \Big\{ s = p + \sum_{k \in \mathbb{Z}} a[k]\rho_{\boldsymbol{\alpha}}(\cdot - kh) \tag{21}$$
$$: a \in \ell_1(\mathbb{Z}), p \in \mathcal{N}_{\boldsymbol{\alpha}} \Big\} \subset \mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}}}(\mathbb{R}).$$

The choice of the search space $\mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$ is obviously guided by Theorem 1, which states that Problem (10) has L-splines solutions, although their knots are not on a uniform grid. The search space therefore contains functions which are close approximations of a solution when the grid size h is small. A mathematical justification is given in Section VI-A.

B. B-Spline Representation and Conditioning

The key property of the search space $\mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$ is that it has an alternative representation in the B-spline basis. This is a fundamental property of cardinal (*i.e.*, $h = 1$) L-splines in general, including cardinal exponential splines [31, Section III. C]. Here, we show the equivalence of these bases for any $h > 0$ in our particular search space $\mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$.

Proposition 1. *The search space $\mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$ can be represented in the B-spline basis as*

$$\mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R}) = \left\{ s = \sum_{k\in\mathbb{Z}} c[k]\beta_{\boldsymbol{\alpha},h}(\cdot - kh) : c \in \ell_{1,h\boldsymbol{\alpha}}(\mathbb{Z}) \right\} \tag{22}$$

where $\beta_{\boldsymbol{\alpha},h}$ is defined as in (17) *and*

$$\ell_{1,h\boldsymbol{\alpha}}(\mathbb{Z}) = \Big\{ (c[k])_{k\in\mathbb{Z}} : (d_{h\boldsymbol{\alpha}} * c) \in \ell_1(\mathbb{Z}) \Big\}. \tag{23}$$

The proof is given in Appendix A. The major contribution of this paper is the use of the B-spline representation (22) of $\mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$ throughout the discretization process, whereas [17] uses the Green's function representation (21). The point of doing so is that contrary to the Green's function, B-splines have finite support; in fact, they are the members of $\mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$ that have minimal support. This makes the B-spline basis close to being orthogonal (it is a Riesz basis [31, Theorem 1]), which leads to Problem (5) being well-conditioned.

The conditioning of an inverse problem is a measure of its numerical stability: a problem is well-conditioned if a small perturbation of the signal coefficients leads to a small perturbation of its measurements. When the basis functions have limited support as in the B-spline case, it is clear that a slight disturbance of the basis coefficients does not change the signal - and thus its measurements - significantly. However, this is not the case in the Green's function case, since not only do the basis functions have infinite support, but they are often non vanishing or even increasing as one moves away from the centers, *e.g.*, $x_+^{N_0-1}$ for $\mathrm{L} = \mathrm{D}^{N_0}$. Therefore, a small perturbation of a basis coefficient greatly affects the reconstructed signal everywhere. The measurements are thus greatly impacted and the problem is severely ill-conditioned [17], [35]. This intuition is confirmed in practice: we observe that in identical settings (*i.e.*, same regularization operator L, measurement operator $\boldsymbol{\nu}$, regularization parameter λ and grid size h), the relevant condition number is systematically greater in Problem (4) than in Problem (5). For example, in the experiment shown in Fig. 3, the condition number of the matrix to be inverted is $\mathrm{cond}(\mathbf{H}_{\beta_\mathrm{L}}^T\mathbf{H}_{\beta_\mathrm{L}} + \lambda\mathbf{L}^T\mathbf{L}) = 5.7\times10^4$ using the B-spline formulation, compared to $\mathrm{cond}(\mathbf{H}_{\rho_\mathrm{L}}^T\mathbf{H}_{\rho_\mathrm{L}} + \lambda\mathbf{I}) = 1.5\times10^{12}$ in the Green's function case. This difference of conditioning largely justifies the use of the B-spline representation of $\mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$ rather than its Green's function representation.

V. Exact Discretization in $\mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$

A. Discrete Problem Formulation

Let $h > 0$ and $\mathrm{L}_{\boldsymbol{\alpha}}$ be an ordinary differential operator. In order to discretize Problem (10) in $\mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$, we use the B-spline representation of $s \in \mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$ given in (22)

$$s(x) = \sum_{k\in\mathbb{Z}} c[k]\beta_{\boldsymbol{\alpha},h}(x - kh) \tag{24}$$

where $c \in \ell_{1,h\boldsymbol{\alpha}}(\mathbb{Z})$. Using (19), the sparsity of s in the Green's function basis is given by $\|d_{h\boldsymbol{\alpha}} * c\|_0$, where $\|\cdot\|_0$ is the ℓ_0 "norm" which counts the number of non-zero entries of a sequence or vector. When we feed (24) into the continuous-domain Problem (10), using (19) once again, we get the following discretized optimization problem

$$\mathcal{S}_h = \underset{c\in\ell_{1,h\boldsymbol{\alpha}}(\mathbb{Z})}{\arg\min} \left\| \sum_{k\in\mathbb{Z}} c[k]\boldsymbol{\nu}(\beta_{\boldsymbol{\alpha},h}(\cdot - hk)) - \mathbf{y} \right\|_2^2 + \frac{\lambda}{h^{N_0-1}}\|d_{h\boldsymbol{\alpha}} * c\|_1, \tag{25}$$

where the associated cost function is denoted by $\mathcal{J}_h : \ell_{1,h\boldsymbol{\alpha}}(\mathbb{Z}) \to \mathbb{R}^+$. By adapting Lemma 20 in [5], it can be shown that $\mathcal{S}_h$ is a non-empty weak*-compact subset of $\ell_{1,h\boldsymbol{\alpha}}(\mathbb{Z})$. Note that Problem (25) is *exactly* equivalent to the continuous-domain Problem (10) restricted to the search space $\mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$. This is the key feature of our formulation: the standard approach to discretize an inverse problem is to use an approximate discrete forward model as a surrogate for a continuous model, which leads to discretization errors. This is not the case of our method, in which the discrete forward model is equal to the continuous one; the former is simply restricted to the native space $\mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$. To the best of our knowledge, aside from [17] of which this paper is an extension, no other work in the literature discretizes non-quadratic continuous-domain problems *exactly* by using a dictionary.

B. Finite Problem

In real-world applications, the signal of interest usually has a given finite support, which we can assume to be $I_T = [0, T]$ without loss of generality. Hence, in all that follows, we can assume that the measurement functionals ν_m, be they ordinary function or distributions such as Dirac impulses, are supported in I_T. This assumption is completely inconsequential for signals supported in I_T, but it is necessary to express (25) as a finite-dimensional problem. In this case, only a finite number of B-spline coefficients affect the data fidelity term in (25); we denote by $I = \{i_{\min}, \dots, i_{\max}\} \subset \mathbb{Z}$ the set of their indices and $N = \#I$. Assuming that $T/h \in \mathbb{N}$, we have $i_{\min} = -N_0+1$, $i_{\max} = T/h-1$ and thus $N = T/h+N_0-1$. To make Problem (25) finite, we optimize over the N B-spline coefficients in I, which are denoted by $\mathbf{c} \in \mathbb{R}^N$. By imposing natural boundary conditions for the regularization term, we get the following discrete finite-dimensional problem.

$$S_h = \underset{\mathbf{c}\in\mathbb{R}^N}{\arg\min} \underbrace{\|\mathbf{H}\mathbf{c} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{L}\mathbf{c}\|_1}_{J(\mathbf{c})} \tag{26}$$

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

$$\mathbf{H} = (\mathbf{h}_{i_{\min}}, \dots, \mathbf{h}_{i_{\max}}) : \quad \mathbf{h}_k = \boldsymbol{\nu}(\beta_{\boldsymbol{\alpha},h}(\cdot - hk)) \tag{27}$$

and the Toeplitz-like regularization matrix $\mathbf{L} : \mathbb{R}^N \to \mathbb{R}^{N-N_0}$ is a finite section of the infinite-dimensional regularization matrix in Problem (25), *i.e.*,

$$\mathbf{L} = \frac{1}{h^{N_0-1}} \begin{pmatrix} d_{h\boldsymbol{\alpha}}[N_0] & \cdots & d_{h\boldsymbol{\alpha}}[0] & 0 & \cdots & 0 \\ 0 & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 \\ 0 & \cdots & 0 & d_{h\boldsymbol{\alpha}}[N_0] & \cdots & d_{h\boldsymbol{\alpha}}[0] \end{pmatrix}. \tag{28}$$

Despite the seemingly arbitrary boundary conditions on the regularization term (we are representing a convolution involving an infinite sequence as a finite matrix multiplication), remarkably, thanks to the finite support assumption, Problem (26) is exactly equivalent to the infinite Problem (25):

Proposition 2. *Under the assumptions of Theorem 1 and assuming that the ν_m functionals are supported in I_T:*

- *We have* $\ker \mathbf{H} \cap \ker \mathbf{L} = \{\mathbf{0}\}$*, and the solution set S_h of Problem* (26) *is a non-empty compact convex set;*
- *Problems* (25) *and* (26) *are equivalent, and there exists a natural bijection between their solutions sets which maps any sequence $c^* \in \mathcal{S}_h \subset \ell_{1,h\boldsymbol{\alpha}}(\mathbb{Z})$ to a vector $\mathbf{c}^* \in S_h \subset \mathbb{R}^N$ such that $c^*|_I = (c^*[i_{\min}], \dots, c^*[i_{\max}]) = \mathbf{c}^*$.*

The proof of Proposition 2 is given in Appendix B. We refer to Problem (26) as being *well-posed* due to the property $\ker \mathbf{H} \cap \ker \mathbf{L} = \{\mathbf{0}\}$, which implies that its solution set S_h is bounded, but not necessarily unique. Proposition 2 demonstrates that the finite Problem (26) is equivalent to the continuous-domain Problem (10) restricted to the search space $\mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$. This is quite a remarkable outcome: we are able to solve an infinite continuous-domain problem in an exact way as a standard discrete inverse problem with ℓ_1 regularization. Once again, this is to the best of our knowledge a novelty, which is this time not present in [17].

Despite the proven equivalence between problems (25) and (26), one might wonder how to proceed in practice to reconstruct the underlying continuous-domain signal s once a solution $\mathbf{c}^* \in S_h$ is reached. By Proposition 2, there exists a unique sequence $c^* \in \mathcal{S}_h$ such $c^*|_I = \mathbf{c}^*$. The following observations, which are direct consequences of the proof of Proposition 2, can be made concerning s:

- The continuous-domain reconstructed signal is $s = \sum_{k\in\mathbb{Z}} c^*[k]\beta_{\boldsymbol{\alpha},h}(\cdot - kh)$;
- The N B-spline coefficients $\mathbf{c}^*$ are sufficient to reconstruct s *exactly* in the interval of interest I_T;
- The sparsity of s in the Green's function basis is given by $\|\mathbf{L}\mathbf{c}^*\|_0$.

These observations indicate that all the relevant information concerning the reconstructed signal s (*i.e.*, its expression in I_T and sparsity) is directly encoded in the vector $\mathbf{c}^* \in S_h$ of Problem (26). Hence, computing the corresponding infinite sequence $c^* \in \mathcal{S}_h$ is unnecessary. In practice, the infinite Problem (25) can thus be altogether forsaken in favor of the computationally feasible finite Problem (26).

C. Reaching a Sparse Solution

In this section, we study the so-called Penalized Basis Pursuit (PBP) problem formulated in (26)

$$S_h = \underset{\mathbf{c}\in\mathbb{R}^N}{\arg\min} \|\mathbf{H}\mathbf{c} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{L}\mathbf{c}\|_1 \tag{29}$$

where the system and regularization matrices $\mathbf{H}$ and $\mathbf{L}$ are defined in (27) and (28) respectively. This problem is close to typical compressed sensing problems ($\mathbf{L}$ is a TV-like regularization matrix), which have been studied at length in the litterature [3], [5] and are known to yield sparse solutions in a certain basis. The specificity of this problem lies in the fact that $\mathbf{L}$ is not invertible. However, Theorem 1 strongly suggests that (26) has sparse solutions, since it is a discretized version of the continuous-domain problem. This instinct is confirmed by the following representer theorem, the proof of which is given in Appendix C.

Theorem 2 (Discrete Representer Theorem)**.** *Let $0 \leq N_0 \leq M < N$, $\mathbf{H} : \mathbb{R}^N \to \mathbb{R}^M$ and $\mathbf{L} : \mathbb{R}^N \to \mathbb{R}^{N-N_0}$ such that* $\ker \mathbf{H} \cap \ker \mathbf{L} = \{\mathbf{0}\}$ *and $\mathbf{L}$ is of full rank,* i.e.*,* $\operatorname{ran} \mathbf{L} = N - N_0$*. Then the solution set S of the optimization problem*

$$S = \underset{\mathbf{c}\in\mathbb{R}^N}{\arg\min} \|\mathbf{H}\mathbf{c} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{L}\mathbf{c}\|_1 \tag{30}$$

is a compact convex set whose extreme points $\mathbf{c}^*$ verify $\|\mathbf{L}\mathbf{c}^*\|_0 \leq M - N_0$.

Theorem 2 is a generalization of Theorem 6 in [5], since it allows for more general regularization matrices (L must be right-invertible). It is also similar to Theorem 2.4 in [16], but with a tighter bound on the sparsity, and with an elementary proof using only standard linear algebra. This result directly applies to Problem (26), since $\ker \mathbf{H} \cap \ker \mathbf{L} = \{\mathbf{0}\}$ by Proposition 2 and $\mathbf{L}$ in (28) is of full rank. Remarkably, the bound on the sparsity $M - N_0$ is the same as for the continuous-domain Problem (32) (Theorem 1), which confirms the close connection between both problems. This is not the case in the Green's function formulation of [17], where the sparsity is bounded by M.

Although Theorem 2 guarantees that Problem (26) has sparse solutions, only the extreme points of S_h are known to be sparse, and in general, S_h is non unique. Therefore, while a solution of Problem (26) can readily be reached using standard solvers such as ADMM, there is no guarantee that this solution will be sparse. In fact, we will demonstrate experimentally later on that, as observed with FISTA in [17], ADMM often converges towards non-sparse solutions, *i.e.*, vectors $\mathbf{c}^*$ such that $\|\mathbf{L}\mathbf{c}^*\|_0 > M - N_0$. To circumvent this issue, we use the following lemma, which is well known in the absence of a regularization matrix [36, Lemma 1]. As it turns out, the latter does not make it more challenging: an elementary proof is given in Appendix D for the sake of completeness.

Lemma 1. *Let $\mathbf{H} : \mathbb{R}^N \to \mathbb{R}^M$, $\mathbf{L} : \mathbb{R}^N \to \mathbb{R}^P$, $\mathbf{y} \in \mathbb{R}^M$ and $\lambda > 0$. We assume that the problem is well posed,* i.e.*,* $\ker \mathbf{H} \cap \ker \mathbf{L} = \{\mathbf{0}\}$*. Then the solution set S_h of the PBP Problem* (26) *is a compact convex set which has a unique measurement*

$\mathbf{y}_\lambda \in \mathbb{R}^M$ *such that* $\forall \mathbf{c} \in S_h$, $\mathbf{Hc} = \mathbf{y}_\lambda$. *Moreover, for any two solutions* $\mathbf{c}_1, \mathbf{c}_2 \in S_h$, *we have*

$$(\mathbf{Lc}_1)_i \times (\mathbf{Lc}_2)_i \geq 0 \quad \forall i \in \{1, \dots, P\}. \tag{31}$$

Note that in the case of Problem (26), as shown in Proposition 2, we have $\ker \mathbf{H} \cap \ker \mathbf{L} = \{\mathbf{0}\}$, which implies that Lemma 1 applies to our problem for $P = N - N_0$. Lemma 1 provides an indirect way of reaching an extreme point of S_h: any solution $\mathbf{c}^* \in S_h$ has a fixed measurement $\mathbf{Hc}^* = \mathbf{y}_\lambda$. Therefore, Problem (26) can be recast as a constrained optimization problem

$$\underset{\mathbf{c} \in \mathbb{R}^N}{\arg\min} \|\mathbf{Lc}\|_1 \qquad s.t. \qquad \mathbf{Hc} = \mathbf{y}_\lambda, \tag{32}$$

which clearly has the same solution set S_h, since the constraint is satisfied for any $\mathbf{c} \in S_h$. This constrained problem can, in turn, be recast as a linear program by introducing the slack variable $\mathbf{u} \in \mathbb{R}^{N-N_0}$:

$$S_h^{\mathrm{LP}} = \underset{(\mathbf{c},\mathbf{u}) \in \mathbb{R}^{2N-N_0}}{\arg\min} \sum_{i=1}^{N-N_0} u_i \qquad s.t. \tag{33}$$

$$\mathbf{u} + \mathbf{Lc} \geq \mathbf{0}; \quad \mathbf{u} - \mathbf{Lc} \geq \mathbf{0}; \quad \mathbf{Hc} = \mathbf{y}_\lambda, \tag{34}$$

where $(\mathbf{c}, \mathbf{u})$ is the concatenation of the vectors $\mathbf{c} \in \mathbb{R}^N$ and $\mathbf{u} \in \mathbb{R}^{N-N_0}$. The following proposition characterizes its solution set S_h^{LP} in terms of S_h:

Proposition 3. S_h^{LP} *is a compact convex set which has extreme points* $(\mathbf{c}^*, \mathbf{u}^*)$ *such that* $\mathbf{c}^*$ *is an extreme point of* S_h.

The proof is given in Appendix E. Proposition 3 allows us to apply the well-known simplex or dual-simplex algorithms [37], [38] to the linear program (33). These algorithms are known to converge to an extreme point $(\mathbf{c}^*, \mathbf{u}^*)$ of the solution set S_h^{LP}. Since $\mathbf{c}^*$ is an extreme point of S_h, Theorem 2 then ensures that it is a sparse solution we are looking for. However, to run this linear program, $\mathbf{y}_\lambda$ needs to be known: hence, we must find a solution (though not necessarily an extreme point) of the PBP Problem (26) beforehand using ADMM or any other suitable algorithm. This solution $\mathbf{c}_{\mathrm{ADMM}} \in S_h$ is then used to compute $\mathbf{y}_\lambda = \mathbf{Hc}_{\mathrm{ADMM}}$, which is used in turn to run the simplex algorithm. This procedure is adapted from [17], in which the same idea is used in the Green's function basis.

VI. Refining the Grid

In the previous sections, we have established an experimental pipeline to solve the continuous-domain problem in $\mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$ for a fixed grid size h. We now study the behavior of the solutions when the grid size h decreases, and how they relate to solutions of the full continuous-domain Problem (10).

A. Convergence of the Cost Function

We place ourselves in the conditions of Theorem 1, which states that there exists at least one solution to the continuous-domain Problem (10) of the form $s(x) = p(x) + \sum_{k=1}^K a_k \rho_{\boldsymbol{\alpha}}(x - x_k)$. This solution does not a priori have knots on a uniform grid, and is thus not included in $\mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$. However, by picking h sufficiently small, it can be approached arbitrarily closely by

$$s_h(x) = p(x) + \sum_{k=1}^K a_k \rho_{\boldsymbol{\alpha}}(x - x_k^h) \in \mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R}), \tag{35}$$

where $x_k^h \in h\mathbb{Z}$ converges to x_k.

The following lemma shows that s_h is indeed a good approximation of s in terms of cost:

Lemma 2. *Let all the hypotheses of Theorem 1 be met for a spline-admissible operator* L. *Then, there exists a family of functions of the form* $s_h = p + \sum_{k=1}^K a_k \rho_{\mathrm{L}}(\cdot - x_k^h) \in \mathcal{M}_{\mathrm{L},h}(\mathbb{R})$ *where* $p \in \mathcal{N}_{\mathrm{L}}$, $K \leq M - N_0$ *and* $x_k^h \in h\mathbb{Z}$ *for any* $h > 0$ *such that*

$$\lim_{h \to 0} \mathcal{J}(s_h) = \min_{f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R})} \mathcal{J}(f) = \mathcal{J}^0. \tag{36}$$

The proof is given in Appendix F. Note that Lemma 2 applies in the most general setting of Theorem 1 and is therefore not limited to differential operators of the form (12).

Going back to exponential B-splines, let $\mathcal{J}_h^0 = \min_{c \in \ell_{1,h\boldsymbol{\alpha}}(\mathbb{Z})} \mathcal{J}_h(c)$ be the optimal cost of the discrete Problem (25). We derive the following theorem, which stems directly from Lemma 2 and is similar to Lemma 8 in [14]:

Theorem 3 (Convergence of the cost function of the discrete problem)**.** *Let all the hypotheses of Theorem 1 be met, and* $\mathrm{L}_{\boldsymbol{\alpha}}$ *be an ordinary differential operator. Then*

$$\lim_{h \to 0} \mathcal{J}_h^0 = \mathcal{J}^0. \tag{37}$$

Proof. Firstly, we observe that

$$\mathcal{J}_h^0 \geq \mathcal{J}^0 \tag{38}$$

since for any $c \in \ell_{1,h\boldsymbol{\alpha}}(\mathbb{Z})$, we can define $s(x) = \sum_{k \in \mathbb{Z}} c[k] \beta_{\boldsymbol{\alpha},h}(x - hk) \in \mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$ which verifies $\mathcal{J}_h(c) = \mathcal{J}(s) \geq \mathcal{J}^0$. Next, let s_h be a family of functions for any $h > 0$ as specified by Lemma 2. Since $s_h \in \mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$, by (22), s_h can be expressed in the B-spline basis as $s_h(x) = \sum_{k \in \mathbb{Z}} c[k] \beta_{\boldsymbol{\alpha},h}(x - hk)$ where $c \in \ell_{1,h\boldsymbol{\alpha}}(\mathbb{Z})$. Therefore, we have

$$\mathcal{J}_h^0 \leq \mathcal{J}_h(c) = \mathcal{J}(s_h) \overset{h \to 0}{\to} \mathcal{J}^0$$

which together with (38) proves the desired result. □

This result can in fact be extended to any operator L which has an admissible B-spline basis. Theorem 3 shows that the choice of $\mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$ as a search space for the continuous-domain problem is a sound one: by solving the discrete problem, we recover a solution which is arbitrarily close in terms of cost to the solution(s) of the continuous problem if h is sufficiently small. Moreover, note that there is no requirement in Theorem 3 that the natural gridded approximation s_h defined in (35) is a solution of the discrete Problem (25): $\mathcal{J}_h^0$ might actually be smaller than $\mathcal{J}(s_h)$. We can therefore hope for a faster convergence than that of $\mathcal{J}(s_h) \to \mathcal{J}^0$.

Input: $\boldsymbol{\nu}$, $\boldsymbol{\alpha}$, T, $\mathbf{y}$, λ, $i_{\min}$, ϵ
Output: $\mathbf{c}^*$
$i = i_{\min}$; $\mathbf{c} = \mathbf{0}$; $cost_p = +\infty$; $error = \epsilon + 1$;
while $error > \epsilon$ **do**
  $h = T/2^i$;
  update $\mathbf{H}$, $\mathbf{L}$; // Depend on $h, \boldsymbol{\nu}, \boldsymbol{\alpha}, T$
  $\mathbf{c} = \mathrm{ADMM}(\mathbf{c}_{\uparrow 2}; \mathbf{H}, \mathbf{L}, \mathbf{y}, \lambda)$;
  $error = |cost(\mathbf{c}) - cost_p|/cost_p$;
  $cost_p = cost(\mathbf{c})$;
  $i = i + 1$;
end
$\mathbf{y}_\lambda = \mathbf{H}\mathbf{c}$;
$\mathbf{c}^* = \mathrm{Simplex}(\mathbf{H}, \mathbf{L}, \mathbf{y}, \lambda, \mathbf{y}_\lambda)$;

Algorithm 1: Pseudocode of our algorithm

B. Multiresolution Strategy

Although Section V-C provides an experimental pipeline to solve the continuous-domain problem in $\mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$ for a fixed grid size h, the choice of the latter is somewhat arbitrary. In practice, in order to choose the grid size, we use the convergence results of Theorem 3. We recursively split the grid in half by taking $h_i = T/2^i$ for increasing values of $i \in \mathbb{N}$, and we solve the corresponding finite problems. This way, the finest grid (highest value of i) contains all its coarser predecessors, which implies that the search spaces are embedded, *i.e.*, $\mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h_i}(\mathbb{R}) \subset \mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h_{i+1}}(\mathbb{R})$. This allows us to use the solution obtained with the previous grid as a starting point of ADMM, which leads to considerable time gains. Another consequence of this embedding is that $\mathcal{J}^0_{h_i} \geq \mathcal{J}^0_{h_{i+1}}$, which indicates that splitting the grid in half can only improve the solution in terms of cost. Theorem 3 then guarantees that $\lim_{i \to +\infty} \mathcal{J}^0_{h_i} = \mathcal{J}^0$. This gives us a natural stopping criterion: we increment i until the relative decrease of cost $(\mathcal{J}^0_{h_{i-1}} - \mathcal{J}^0_{h_i})/\mathcal{J}^0_{h_{i-1}}$ is smaller than some tolerance parameter ϵ. When ϵ is sufficiently small, we consider that the cost function has converged and that there is no need to make the grid any finer. Note that the simplex step is only necessary for the final grid size. This complete procedure is detailed in Algorithm 1.

In Algorithm 1, $\mathrm{ADMM}(\mathbf{c}_{\uparrow 2}; \mathbf{H}, \mathbf{L}, \mathbf{y}, \lambda)$ runs ADMM on Problem (26) with the starting point $\mathbf{c}_{\uparrow 2}$. The latter corresponds to the vector of B-spline coefficients $\mathbf{c}$ converted to the current grid size, which is twice as fine as that of $\mathbf{c}$. This conversion is made possible by the embedding of the search spaces. Similarly, $\mathrm{Simplex}(\mathbf{H}, \mathbf{L}, \mathbf{y}, \lambda, \mathbf{y}_\lambda)$ runs the simplex algorithm on the constrained Problem (33) (no starting point is required). The output $\mathbf{c}^*$ of this algorithm is therefore a vector whose size is not predetermined, but which represents a continuous-domain signal in I_T that is sparse in the Green's function basis, and yields a cost close to $\mathcal{J}_0$.

VII. Numerical Experiments

We now discuss our implementation of Algorithm 1 and present some results. The differential regularization operators that we consider in our experiments either have identical or unique poles. Our algorithms are implemented using GlobalBioIm [39], an inverse problem library developed in our group, as well as the Gurobi optimizer [40] for the simplex algorithm.

A. Experimental Setting

1) Test signal: We attempt to reconstruct sparse signals of the form

$$s(x) = \sum_{k=1}^{K_s} a_k \rho_{\mathrm{L}}(x - x_k) + \sum_{n=1}^{N_0} b_n p_n(x), \qquad (39)$$

for which gTV regularization is an adequate prior by Theorem 1. The sparsity index K_s is chosen by the user and the knots x_k are drawn at random in the interval of interest I_T following a uniform distribution. The coefficients a_k and b_n are i.i.d Gaussian random variables projected on the subspace of vectors $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^{K+N_0}$ for which s is supported in I_T. This is to enforce the finite support assumption on the test signal, which is implicit in the discrete problem formulation (26). Therefore, aside from the approximation error on the knot locations, the test signal in (39) is in the span of feasible signals reconstructed by the discrete problem, which is obviously a desirable property.

2) Measurements: We implemented three types of measurement operators $\boldsymbol{\nu}$:

- **Ideal sampling:** This case corresponds to a measurement operator $\boldsymbol{\nu}(f) = \Big(\langle \delta(\cdot - x_1), f\rangle, \ldots, \langle \delta(\cdot - x_M), f\rangle\Big) = \Big(f(x_1), \ldots, f(x_M)\Big)$, where $\mathbf{x} \in (I_T)^M$. Given the form of $\mathcal{N}_{\mathrm{L}}$ in (13), it can be shown that $\boldsymbol{\nu}$ satisfies the well-posedness assumption $\mathcal{N}_{\boldsymbol{\nu}} \cap \mathcal{N}_{\mathrm{L}} = \{0\}$ as soon as $M \geq N_0$ and all sampling points x_m are pairwise distinct. We either take uniformly spaced knots or random samples following a uniform distribution in I_T for $\mathbf{x}$.
- **Fourier sampling:** For an odd number of measurements M, we define the measurement operator $\boldsymbol{\nu}(f) = \Big(\widehat{f}_{|I_T}(0), \mathrm{Re}(\widehat{f}_{|I_T}(\omega_2)), \mathrm{Im}(\widehat{f}_{|I_T}(\omega_2)), \ldots, \mathrm{Re}(\widehat{f}_{|I_T}(\omega_{\frac{M+1}{2}})), \mathrm{Im}(\widehat{f}_{|I_T}(\omega_{\frac{M+1}{2}}))\Big)$ where $f_{|I_T} = f \times \mathbb{1}_{[0,T]}$ is f to which a rectangular window function on the interval I_T is applied. The sampling frequencies are $\omega_1 = 0$ and ω_m for $2 \leq m \leq M$ which are drawn from a uniform distribution in $(0, \omega_{\max}]$. Since $\widehat{s}(0) \in \mathbb{R}$ when s is a real signal, there is no need to take the real and imaginary parts; however, this is necessary for non-zero frequencies in order to have real measurements. The maximum frequency $\omega_{\max}$ is chosen such that the spectrum of s has small energy above this threshold.
- **Inner product with dephased cosines:** This type of measurement is a variation of Fourier sampling: we take the inner product with functions $x \mapsto \cos(\omega_m x + \theta_m)$ where the θ_m are drawn from a uniform distribution in $[0, \pi)$. Compared to Fourier sampling where each sampling frequency accounts for two measurements, this allows for the sampling of a broader spectrum of frequencies: $\boldsymbol{\nu}(f) = \Big(\widehat{f}_{|I_T}(0), \langle \cos(\omega_2 x + \theta_2), f_{|I_T}\rangle, \ldots, \langle \cos(\omega_M x + \theta_M), f_{|I_T}\rangle\Big)$, where as for Fourier sampling, $\omega_1 = 0$ and $\omega_m \in (0, \omega_{\max}]$ for $m = \{2, \ldots, M\}$.

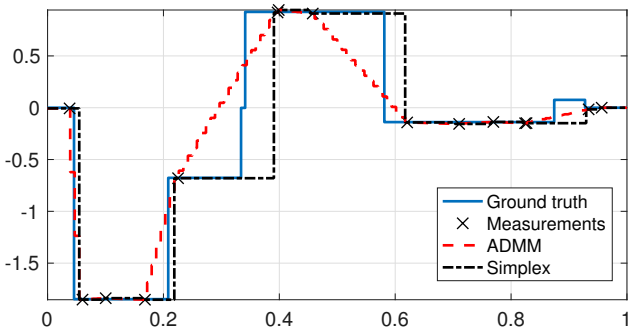


Fig. 1: Reconstruction results for $\mathrm{L} = \mathrm{D}$, $M = 15$ (ideal sampling), grid size $h = \frac{1}{128}$.
Sparsity: 118 after ADMM, 13 after simplex.

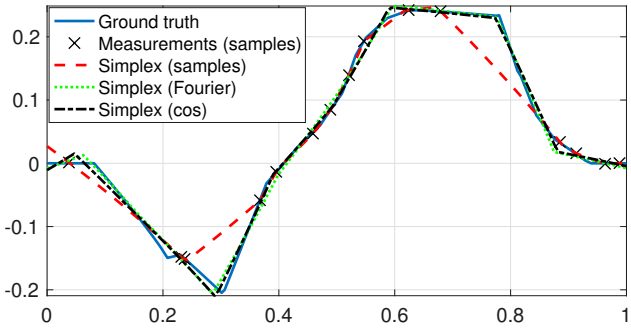


Fig. 2: Reconstruction results for $\mathrm{L} = \mathrm{D}(\mathrm{D} - \mathrm{I})$, $M = 15$ measurements, grid size $h = \frac{1}{128}$.
Ideal sampling $(--)$: $\lambda = 7.71 \cdot 10^{-9}$, Sparsity$= 13$, SNR$=$ 13.76 dB;
Fourier sampling $(\cdot\cdot)$: $\lambda = 1.35 \cdot 10^{-8}$, Sparsity$= 12$, SNR$=$ 24.04 dB;
Dephased cosine sampling $(-\cdot)$: $\lambda = 6.12 \cdot 10^{-9}$, Sparsity$=$ 13, SNR$= 24.90$ dB.

We compare experimental results using these measurement operators, predicting that reconstructed signals should be closer to the test signals when sampling in the Fourier domain than with ideal sampling. This ensues from the theory of compressed sensing: Fourier matrices are known to have good recovery properties with few measurements [41], whereas sampling matrices clearly do not. In order to be more realistic and to verify the robustness of our algorithm, we add Gaussian noise to the measurements with standard deviation σ computed from a given Signal-to-Noise Ratio (SNR).

3) Regularization Parameter: The choice of the regularization parameter λ is critical, as it greatly affects the reconstructed signal: high values of λ can lead to overly regularized solutions, whereas low values tend to suppress the effect of regularization. The value of λ should be tuned according to the type of measurement. To this end, in our experiments, we choose a value of λ among a list of potential values such that the SNR between the reconstructed signal and the test signal is the highest for a certain value of h. The selected λ is then used for all values of h, as specified in Algorithm 1.

B. Experimental Results

We now present several results of our numerical implementation.

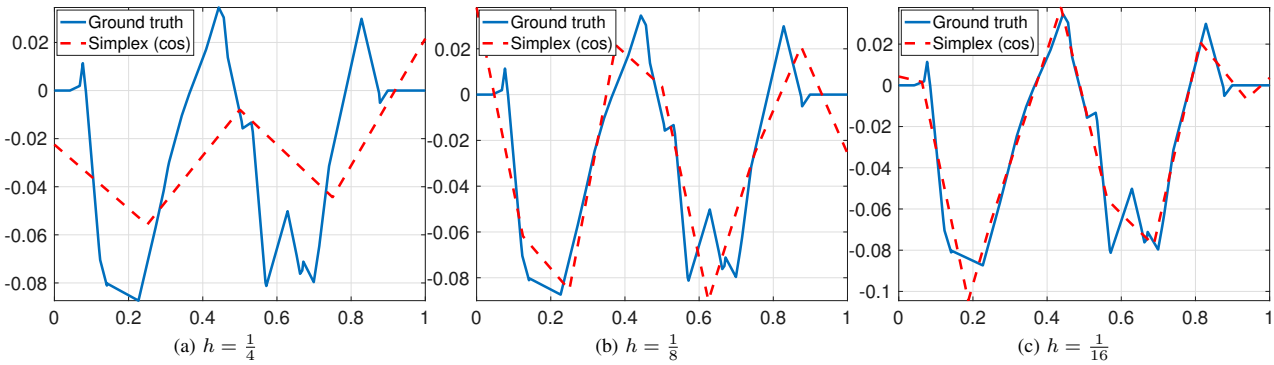
1) Sparsity: In our experiments, we observe that, as predicted by Theorem 2, the final reconstructed signal has sparsity $K \leq M - N_0$ in the Green's function basis.

However, running the simplex after ADMM is far from being superfluous: reconstructed signals after ADMM are typically not sparse at all. This is best illustrated in the case of ideal sampling, where we observe a staircase effect between measurements (Fig. 1). Although this phenomenon does not affect the cost function, it is clearly not optimal in terms of sparsity, and it illustrates the non-uniqueness of S_h. However, after the simplex step, the sparsity improves dramatically, going from 118 to $13 \leq M - N_0$ in Fig. 1, as predicted by Theorem 2.

2) Measurement Types Comparison: Fig. 2 shows a typical example of reconstruction results after the simplex step for all three measurement types, with an identical grid size $h = 1/128$. The SNR values are computed with respect to the test signal; they show that as predicted, reconstruction results are much better in terms of SNR when sampling is done in the Fourier domain. Indeed, for ideal sampling, when there are no measurement points in the vicinity of a jump (around $x = 0.3$ and $x = 0.8$ in Fig. 2), the reconstruction result can vastly deviate from the test signal. On the other hand, the reconstruction using Fourier and dephased cosine samples is remarkably similar to the test signal, despite the small number of measurements ($M = 15$) and a test signal with comparatively high sparsity ($K_s = 30$). Note that in all three cases, the reconstructed signals have sparsity $K \leq M - N_0 = 13$ which conforms with Theorem 2.

3) Decreasing Grid Size: As the grid size decreases, the search space of our optimization problem becomes larger: we can therefore reconstruct functions in finer detail. This is illustrated in Fig. 3, in which we observe that very coarse grids approximate complex signals very poorly, whereas after splitting the grid in half recursively, these signals can rapidly be approximated much better (Fig. 3c).

To illustrate the effect of the decreasing grid size in terms of cost, we present an example run of Algorithm 1 with a regularization operator $\mathrm{L} = \mathrm{D}^3$ in Fig. 4. The final reconstructed signal is shown in Fig. 4a: notwithstanding the reasonably fine grid size ($h = 1/2^8$), the reconstruction is near-perfect. The evolution of the cost function with respect to the grid size in our example is shown in Fig. 4b: we observe that after an initial rapid decrease, the cost function starts plateauing, which is in line with Theorem 3. Given the aspect of this evolution, it is safe to assert that the cost is very close to its limit value $\mathcal{J}_0$ specified by Theorem 3. Although we could consider tightening the tolerance ϵ to get a marginally smaller cost, this is not necessarily a sensible choice. Indeed, for very fine grids (*e.g.*, $h < 1/2^{11}$), the increased scale of the problem can cause computational problems larger than the potential gain in terms of cost. We found the choice of $\epsilon = 10^{-3}$ to be a good compromise in our experimental setting: the



(a) $h = \frac{1}{4}$ (b) $h = \frac{1}{8}$ (c) $h = \frac{1}{16}$

Fig. 3: Reconstruction results after simplex for $\mathrm{L} = \mathrm{D}^2$, $M = 31$ (dephased cosine sampling) with decreasing h.

final grid size h is typically coarser than $1/2^{10}$, even for very non-sparse test signals ($K_s \approx 100$) and with many measurements ($M \approx 100$). For such grid sizes, due to the good conditioning of the system matrix $\mathbf{H}$, the optimization problems are entirely feasible (ADMM typically converges in a few seconds with a properly tuned penalty parameter ρ) and computational problems are avoided.

In order to compare reconstruction results for different grid sizes, we applied the simplex step as described in Sec. V-C for every grid size h in our example. Despite the convergence of the cost function for the finer grid sizes observed in Fig. 4b, the variations in the sparsity (Fig. 4c) indicate that the reconstructed signals are not identical from one grid size to the next. However, in regard to the optimization problem, there is no reason to decrease the grid size any further or to favor one solution over another if both yield the same cost within a user-defined tolerance.

C. Comparison with Discrete Methods

In this section, we assess the pertinence of our framework by comparing it with a purely discrete method. The standard way of discretizing Problem (10) would be to consider uniform samples of the reconstructed function, *i.e.*, a pixel basis, and to approximate derivative operators with finite differences. Within this framework, the underlying discrete optimization problem of the form (26) is very similar for both methods. Indeed, in both cases, the regularization matrix $\mathbf{L}$ is a finite difference-type matrix as in (28). However, since the basis functions are different, the number of coefficients N and the system matrix $\mathbf{H}$ differ. We solve both problems using our pipeline described in V-C.

We consider noiseless Fourier-domain measurements (dephased cosine sampling) with $M = 100$. As explained earlier, this bolsters the recovery properties of the reconstruction, and thus allows us to use similarity metrics between the reconstructed and test signals to compare both methods.

Such a comparison is made in Figure 5, with $\mathrm{L} = \mathrm{D}^4$ and $\lambda = 10^{-15}$. For the sake of fairness, we use a piecewise linear test signal, since the latter does not resemble the basis functions of either method. Figure 5a shows the reconstruction result, using a coarse grid for visualization purposes. Our continuous-domain reconstruction is clearly a lot closer to the test signal; this observation is confirmed by looking at the SNR of both reconstructions in Figure 5b. We notice that the SNR is similar for both methods using finer grids: this is in keeping with [42], which demonstrates some form of convergence of discrete methods towards solutions of continuous-domain problems as the grid size goes to zero. However, our continuous-domain method converges much faster (*i.e.*, for coarser grids) towards a very faithful reconstruction.

Note that the observed linear regime of the blue curve in Fig. 5b is consistent with the approximation power of pixels, which is well known to be in $\mathcal{O}(h)$. Moreover, using finite differences instead of the derivative yields additional errors which increase with the order N_0 of the operator and when the grid gets coarser. Conversely, our method is exact in the continuous domain for *any* grid size, which explains why a grid of $h = 1/2^5$ can be sufficiently fine. Finally, note that although the discrete method leads to a slightly better-conditioned problem, the difference in speed is negligible due to the Riesz basis property of exponential B-splines.

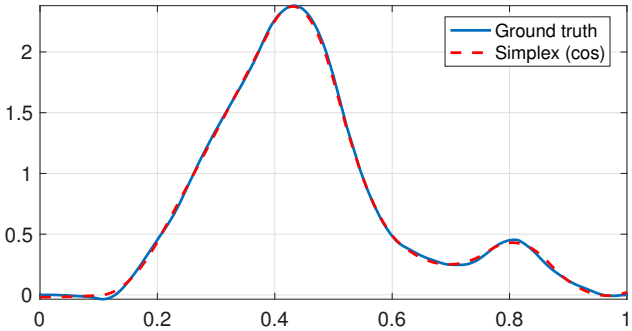
VIII. Conclusion

In this paper, we have devised an efficient multiresolution algorithm to compute sparse solutions of continuous-domain inverse problems with gTV regularization numerically. Our grid-based discretization uses the B-spline dictionary basis matched to the operator L. On the theoretical side, we proved that this is an exact discretization of the underlying continuous-domain problem restricted to a search space, and that this discrete problem converges in terms of cost towards the continuous problem when the grid size decreases. On the experimental side, we implemented this discretization scheme for ordinary differential regularization operators L, and several different measurement operators. Our experimental results demonstrate that our formulation is computationally inexpensive, well suited for practical problems and compares favorably to standard, purely discrete methods.

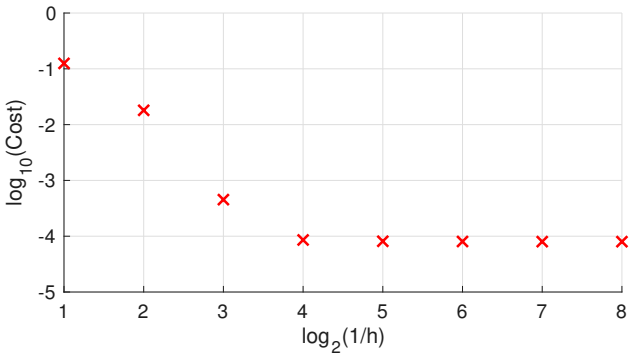
Appendix

A. Proof of Proposition 1

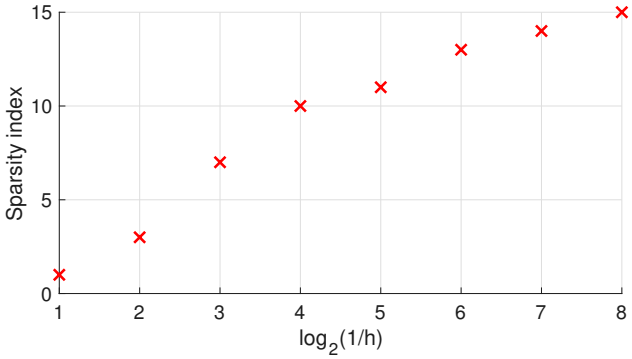
We first prove the inverse inclusion $c \in \ell_{1,h\boldsymbol{\alpha}}(\mathbb{Z}) \Rightarrow s = \sum_{k\in\mathbb{Z}} c[k]\beta_{\boldsymbol{\alpha},h}(\cdot - kh) \in \mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$. The innovation (19) of



(a) Reconstructed signal (grid size $h = \frac{1}{2^8}$).

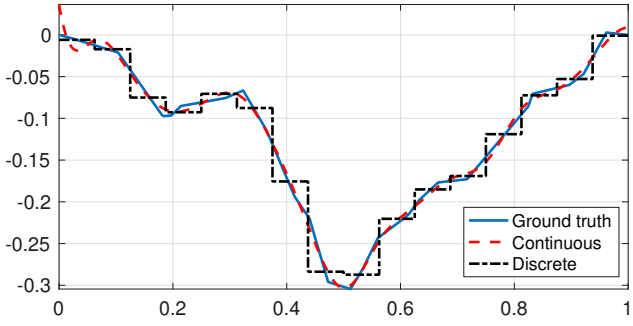


(b) Evolution of the cost of the reconstructed signal.

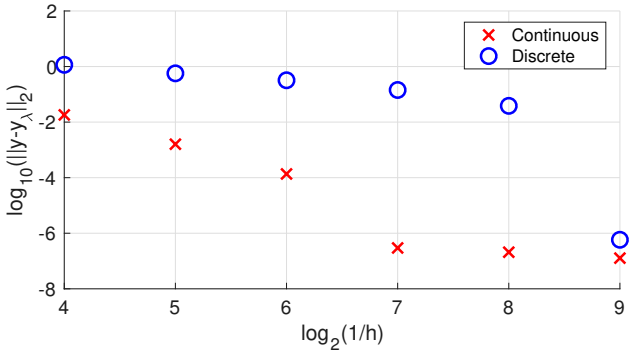


(c) Evolution of the sparsity index of the reconstructed signal.

Fig. 4: Example run for $\mathrm{L} = \mathrm{D}^3$ (quadratic splines), $M = 31$ (dephased cosine sampling).



(a) Reconstructed signals (grid size $h = \frac{1}{16}$).



(b) Evolution of the SNR with respect to the test signal.

Fig. 5: Comparisons between our continuous model and the pixel-based discrete model for $\mathrm{L} = \mathrm{D}^4$, $M = 100$ (dephased cosine sampling).

the exponential B-spline shows that the latter is indeed an L-spline in the sense of Definition 2. Next, a simple calculation using (19) yields $a = \frac{1}{h^{N_0-1}}(d_{h\boldsymbol{\alpha}} * c) \in \ell_1(\mathbb{Z})$, which implies that $s \in \mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$.

We now show the direct inclusion, *i.e.*, that $\mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$ is spanned by B-splines. It can be shown [31, Equation (22)] that there exists a unique sequence $p_{h\boldsymbol{\alpha}}$ that is an inverse of $d_{h\boldsymbol{\alpha}}$ for the discrete convolution product and verifies

$$\rho_{\boldsymbol{\alpha}}(x) = h^{N_0-1} \sum_{k\in\mathbb{Z}} p_{h\boldsymbol{\alpha}}[k]\beta_{\boldsymbol{\alpha},h}(x - hk). \tag{40}$$

This sequence $p_{h\boldsymbol{\alpha}}$ is slowly growing with the same growth rate n_0 as $\rho_{\boldsymbol{\alpha}}$, meaning that $p_{h\boldsymbol{\alpha}}[k]/|k+1|^{n_0}$ is a bounded sequence. Next, the proof that $\mathcal{N}_{\boldsymbol{\alpha}}$ is spanned by the cardinal B-spline basis is given in [31, Section III., C., 2)]. Both these properties are given for cardinal B-splines (*i.e.*, $h = 1$), and they can be adapted without difficulty for B-splines with knots spacing h. There only remains to prove that the sequence c of B-spline coefficients of $s = p + \sum_{k\in\mathbb{Z}} a[k]\rho_{\boldsymbol{\alpha}}(\cdot - kh) \in \mathcal{M}_{\mathrm{L}_{\boldsymbol{\alpha}},h}(\mathbb{R})$ (*i.e.*, $a \in \ell_1(\mathbb{Z})$) is in $\ell_{1,h\boldsymbol{\alpha}}(\mathbb{Z})$. A simple calculation yields $c = h^{N_0-1}(p_{h\boldsymbol{\alpha}} * a)$, which is clearly well defined when $a \in \mathcal{S}(\mathbb{Z})$ (the space of rapidly-decreasing sequences) since $p_{h\boldsymbol{\alpha}}$ is slowly growing. Next, $\mathcal{S}(\mathbb{Z})$ is dense in $\ell_1(\mathbb{Z})$, which allows us to extend this definition to any $a \in \ell_1(\mathbb{Z})$ by continuity [5, Theorem 16]. We thus have $d_{h\boldsymbol{\alpha}} * c = a \in \ell_1(\mathbb{Z})$. Finally, for elements of $\mathcal{N}_{\boldsymbol{\alpha}}$, the sequence c of B-spline coefficients verifies $d_{h\boldsymbol{\alpha}} * c = 0 \in \ell_1(\mathbb{Z})$. This proves the direct inclusion and thus the desired result.

B. Proof of Proposition 2

The second item of Proposition 2 entails the existence of bijective linear map $\theta : S_h \to \mathcal{S}_h$ such that $\theta(\mathbf{c}^*)|_I = \mathbf{c}^*$ for any $\mathbf{c}^* \in S_h$. In order to construct this mapping, we rely on the following lemma:

Lemma 3. *Assume that the ν_m ($1 \le m \le M$) functionals are supported in I_T. Then solutions $c^* \in \mathcal{S}_h$ of Problem* (25) *are uniquely determined by their N coefficients $c^*|_I$.*

Proof. Let $c^* \in \mathcal{S}_h$ be a solution of the discrete Problem (25). Consider a sequence c such that $c|_I = c^*|_I$ and whose remaining coefficients are free. The latter do not affect the

data fidelity term due to the finite support assumption on the ν_m. When $N > N_0$, $c[i_{\max}+1]$ can be uniquely chosen such that $(d_{h\boldsymbol{\alpha}} * c)[i_{\max}+1] = \sum_{k=0}^{N_0} d_{h\boldsymbol{\alpha}}[k]c[i_{\max}-k+1] = 0$. Similarly, all $c[k]$ coefficients for $k > i_{\max}$ can be uniquely determined recursively to nullify $(d_{h\boldsymbol{\alpha}} * c)[k]$ as a linear combination of the $N_0 - 1$ previous coefficients. The same can be done for coefficients $c[k]$ with $k \leq 0$, using this time the $N_0 - 1$ following coefficients of c. By construction, this sequence c yields a regularization cost smaller or equal to that of c^*, and since both yield the same data fidelity cost, we have $\mathcal{J}_h(c) \leq \mathcal{J}_h(c^*)$. Since c^* is a solution of (25) and the construction of c is unique, we necessarily have $c = c^*$. $\square$

The proof of Lemma 3 details the construction of an injective linear map $\tilde{\theta} : \mathbb{R}^N \to \ell_{1,h\boldsymbol{\alpha}}(\mathbb{Z})$ such that for any $\mathbf{c} \in \mathbb{R}^N$, $\tilde{\theta}(\mathbf{c})|_I = \mathbf{c}$. Let $\mathbf{c} \in \mathbb{R}^N$, and consider the corresponding sequence $\tilde{\theta}(\mathbf{c}) \in \ell_{1,h\boldsymbol{\alpha}}(\mathbb{Z})$. Following the proof of Lemma 3, $\mathcal{J}_h(\tilde{\theta}(\mathbf{c}))$ can be computed using only the N coefficients $\mathbf{c}$. Indeed, all other coefficients $(\tilde{\theta}(\mathbf{c})[k])_{k \notin I}$ do not affect the data fidelity term and cancel out all the regularization terms which they affect. This implies that $\mathcal{J}_h(\tilde{\theta}(\mathbf{c})) = J_h(\mathbf{c}) = \|\mathbf{H}\mathbf{c}-\mathbf{y}\|_2^2+\lambda\|\mathbf{L}\mathbf{c}\|_1$, where $\mathbf{H}$ and $\mathbf{L}$ are defined as in (27) and (28) respectively. Since by Lemma 3, $\mathcal{S}_h \subset \tilde{\theta}(\mathbb{R}^N)$, problems (25) and (26) are equivalent in the sense that $\tilde{\theta}(S_h) = \mathcal{S}_h$, and the restriction $\theta = \tilde{\theta}|_{S_h} : S_h \to \mathcal{S}_h$ is a bijective linear map.

Concerning the first item of Proposition 2, let $\mathbf{c} \in \ker \mathbf{H} \cap \ker \mathbf{L}$, the corresponding signal s verifies $s = \sum_{k\in\mathbb{Z}} \tilde{\theta}(\mathbf{c})[k]\beta_{\boldsymbol{\alpha},h}(\cdot - kh) \in \mathcal{N}_{\boldsymbol{\nu}} \cap \mathcal{N}_{\boldsymbol{\alpha}} = \{0\}$ (well-posedness assumption in Theorem 1), which implies that $\mathbf{c} = 0$. Hence, $\ker \mathbf{H} \cap \ker \mathbf{L} = \{\mathbf{0}\}$, which implies that Problem (26) is well-posed and thus that its solution set S_h is a non-empty compact set. The latter is also convex due to the convexity of the cost function J_h.

C. Proof of Theorem 2

Let $J : \mathbf{c} \mapsto \|\mathbf{H}\mathbf{c} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{L}\mathbf{c}\|_1$. Since J is continuous and coercive due to the well-posedness assumption $\ker \mathbf{H} \cap \ker \mathbf{L} = \{\mathbf{0}\}$, S is a non-empty, closed compact set. Therefore, by the Krein-Milman theorem, it is the closed convex hull of its extreme points.

Let $\mathbf{c}^*$ be an extreme point of S. Assume by contradiction that $\mathbf{L}\mathbf{c}^*$ has sparsity $K > M - N_0$, i.e. $\mathbf{L}\mathbf{c}^* = \sum_{k=1}^{K} a_{n_k}\mathbf{e}_{n_k}$ where the $n_k \in \{1, \ldots, N\}$ are distinct, $a_{n_k} \neq 0$ and $\{\mathbf{e}_i\}_{i=1}^{N-N_0}$ is the canonical basis of $\mathbb{R}^{N-N_0}$. Consider the vector space $T = \operatorname{ran} \mathbf{L} \cap \operatorname{span}\{\mathbf{e}_{n_k}\}_{k=1}^{K}$. To find a lower bound on the dimension of T, we use the relation

$$\begin{aligned}\dim(X \cap Y) &= \dim X + \dim Y - \dim(X + Y) \\ &\geq \dim X + \dim Y - P, \end{aligned} \tag{41}$$

where X and Y are vector subspaces of $\mathbb{R}^P$. Since the rank of $\mathbf{L}$ is $N - N_0$, for $X = \operatorname{ran} \mathbf{L}$ and $Y = \{\mathbf{e}_{n_k}\}_{k=1}^{K}$, (41) yields $R = \dim T \geq K > M - N_0$ (with $P = N - N_0$). Let $\{\mathbf{t}_r\}_{r=1}^{R}$ be a basis of T. By definition of T, there exist vectors $\mathbf{g}_r$ and coefficients $t_k^r \in \mathbb{R}$ such that $\mathbf{t}_r = \mathbf{L}\mathbf{g}_r = \sum_{k=1}^{K} t_k^r \mathbf{e}_{n_k}$.

Next, we define $\mathbf{y}_r = \mathbf{H}\mathbf{g}_r \in \mathbb{R}^M$ for all $r \in \{1, \ldots, R\}$, and $\mathbf{z}_n = \mathbf{H}\mathbf{p}_n \in \mathbb{R}^M$ for all $n \in \{1, \ldots, N_0\}$ where $\{\mathbf{p}_n\}_{n=1}^{N_0}$ is a basis of $\ker \mathbf{L}$. The collection of vectors $\{\mathbf{y}_1, \ldots, \mathbf{y}_R, \mathbf{z}_1, \ldots, \mathbf{z}_{N_0}\}$ has $R + N_0 \geq K + N_0 > M$ elements, and is thus linearly dependent. Therefore, there exist coefficients $\alpha_r, \beta_n \in \mathbb{R}$ such that $\sum_{r=1}^{R} \alpha_r \mathbf{y}_r + \sum_{n=1}^{N_0} \beta_n \mathbf{z}_n = \mathbf{0}$ and $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \neq \mathbf{0}$. We then define $\mathbf{c}_0 = \sum_{r=1}^{R} \alpha_r \mathbf{g}_r + \sum_{n=1}^{N_0} \beta_n \mathbf{p}_n \in \mathbb{R}^N$, which is clearly in $\ker \mathbf{H}$. Assume by contradiction $\mathbf{L}\mathbf{c}_0 = \sum_{r=1}^{R} \alpha_r \mathbf{t}_r = \mathbf{0}$. We thus have $\mathbf{c}_0 \in \ker \mathbf{H} \cap \ker \mathbf{L} = \{\mathbf{0}\}$. Moreover, since the $\mathbf{t}_r$ are linearly independent, we have $\boldsymbol{\alpha} = \mathbf{0}$, and thus $\mathbf{c}_0 = \sum_{n=1}^{N_0} \beta_n \mathbf{p}_n = \mathbf{0}$. Yet the $\mathbf{p}_n$ are also linearly independant, which means that $\boldsymbol{\beta} = \mathbf{0}$, which contradicts $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \neq \mathbf{0}$. Therefore, we have $\mathbf{L}\mathbf{c}_0 \neq \mathbf{0}$, which implies that $\mathbf{c}_0 \neq \mathbf{0}$.

Finally, we pick an $\epsilon > 0$ such that

$$\epsilon < \frac{\min_k a_{n_k}}{\max_k |\sum_{r=1}^{R} \alpha_r t_k^r|}. \tag{42}$$

Note that ϵ is well defined since for all k, $a_{n_k} > 0$ and $\sum_{r=1}^{R} \alpha_r t_k^r = 0$ for all k would imply that $\mathbf{L}\mathbf{c}_0 = \sum_{r=1}^{R} \alpha_r \mathbf{t}_r = \sum_{k=1}^{K} \left(\sum_{r=1}^{R} \alpha_r t_k^r\right) \mathbf{e}_{n_k} = \mathbf{0}$, which we have proved to be false. We can then compute:

$$\begin{aligned}\|\mathbf{L}(\mathbf{c}^* \pm \epsilon\mathbf{c}_0)\|_1 &= \|\sum_{k=1}^{K} \left(a_{n_k} \pm \epsilon \sum_{r=1}^{R} \alpha_r t_k^r\right) \mathbf{e}_{n_k}\|_1 \\ &= \sum_{k=1}^{K} \left|a_{n_k} \pm \epsilon \sum_{r=1}^{R} \alpha_r t_k^r\right| \\ &= \|\mathbf{L}\mathbf{c}^*\|_1 \pm \epsilon \sum_{k=1}^{K}\sum_{r=1}^{R} \alpha_r t_k^r \end{aligned}$$

since by definition of ϵ, $a_{n_k} \pm \epsilon \left(\sum_{r=1}^{R} \alpha_r t_k^r\right) > 0$ for all k. Notice that both vectors $(\mathbf{c}^* \pm \epsilon\mathbf{c}_0)$ yield the same data fidelity cost as $\mathbf{c}^*$ in Problem (26): indeed, $\mathbf{H}(\mathbf{c}^* \pm \epsilon\mathbf{c}_0) = \mathbf{H}\mathbf{c}^*$ since $\mathbf{c}_0 \in \ker \mathbf{H}$. Therefore, if $\sum_{k=1}^{K}\sum_{r=1}^{R} \alpha_r t_k^r \neq 0$, then either $(\mathbf{c}^* + \epsilon\mathbf{c}_0)$ or $(\mathbf{c}^* - \epsilon\mathbf{c}_0)$ yields a cost strictly smaller than that of $\mathbf{c}^*$ in Problem (26), which is impossible since $\mathbf{c}^*$ is a solution of the latter. Consequently, $\sum_{k=1}^{K}\sum_{r=1}^{R} \alpha_r s_k^r = 0$ and so $(\mathbf{c}^* \pm \epsilon\mathbf{c}_0) \in S$. Yet $\mathbf{c}^* = \frac{1}{2}(\mathbf{c}^* + \epsilon\mathbf{c}_0) + \frac{1}{2}(\mathbf{c}^* - \epsilon\mathbf{c}_0)$, and since $\epsilon\mathbf{c}_0 \neq \mathbf{0}$, $\mathbf{c}^*$ is not an extreme point of S, which contradicts our initial assumption. This proves the desired result $K \leq M - N_0$.

D. Proof of Lemma 1

Due to the well-posedness assumption, the cost function J is coercive and since it is continuous, S_h is non-empty and bounded. Let $\mathbf{c}_1, \mathbf{c}_2 \in S_h$ be two (possibly identical) solutions. We have $J(\mathbf{c}_1) = J(\mathbf{c}_2) = J_0$, and for any $\alpha \in [0, 1]$, we define $\mathbf{c}_\alpha = \alpha\mathbf{c}_1 + (1 - \alpha)\mathbf{c}_2$. The convexity of J yields $J(\mathbf{c}_\alpha) \leq \alpha J(\mathbf{c}_1) + (1 - \alpha)J(\mathbf{c}_2) = J_0$. Yet, since J_0 is the minimum of the cost function J, we have $J(\mathbf{c}_\alpha) = J_0$, which implies that $\mathbf{c}_\alpha \in S_h$ and thus that S_h is a convex set. Another implication is that the convexity inequality is in fact an equality. For the data fidelity term, the strict convexity of the squared ℓ_2 norm implies that $\mathbf{H}\mathbf{c}_1 - \mathbf{y} = \mathbf{H}\mathbf{c}_2 - \mathbf{y} \Leftrightarrow \mathbf{H}\mathbf{c}_1 = \mathbf{H}\mathbf{c}_2 = \mathbf{y}_\lambda$.

The second property (31) results from the case of equality in the triangular inequality of the ℓ_1 norm: we have $\|\mathbf{L}\mathbf{c}_\alpha\|_1 =$

$\alpha\|\mathbf{L}\mathbf{c}_1\|_1 + (1-\alpha)\|\mathbf{L}\mathbf{c}_2\|_1$. Each coordinate can be treated separately, yielding

$$\begin{aligned} &|\,(\mathbf{L}\mathbf{c}_\alpha)_i\,| = \alpha|(\mathbf{L}\mathbf{c}_1)_i| + (1-\alpha)|(\mathbf{L}\mathbf{c}_2)_i| \\ \Leftrightarrow &(\mathbf{L}\mathbf{c}_1)_i \times (\mathbf{L}\mathbf{c}_2)_i \geq 0 \qquad \forall i \in \{1,\dots,P\}. \end{aligned}$$

E. Proof of Proposition 3

We prove the following statement, which is stronger than that of Proposition 3:
S_h^{LP} is a compact convex set, and S_h and S_h^{LP} have corresponding extreme points through the one-to-one mapping $\phi : S_h \to S_h^{\mathrm{LP}}$ defined by $\phi(\mathbf{c}) = (\mathbf{c}, |\mathbf{L}\mathbf{c}|)$ and its inverse $\phi^{-1} : S_h^{\mathrm{LP}} \to S_h$ defined by $\phi^{-1}((\mathbf{c},\mathbf{u})) = \mathbf{c}$.

Proof. Let us first observe that S_h^{LP} is of the form $S_h^{\mathrm{LP}} = \{(\mathbf{c}, |\mathbf{L}\mathbf{c}|) \in \mathbb{R}^{2N-N_0},\ \mathbf{c} \in \mathbb{R}^N\}$, where $|\mathbf{x}|$ is the vector of component-wise absolute values of $\mathbf{x}$. This implies that $\{\mathbf{c} \in \mathbb{R}^N,\ (\mathbf{c}, |\mathbf{L}\mathbf{c}|) \in S_h^{\mathrm{LP}}\}$ is the solution set of the constrained optimization Problem (32), which is equal to S_h. Therefore, we have proved that $S_h = \{\mathbf{c} \in \mathbb{R}^N,\ (\mathbf{c}, |\mathbf{L}\mathbf{c}|) \in S_h^{\mathrm{LP}}\}$, and thus that $S_h^{\mathrm{LP}} = \{(\mathbf{c}, |\mathbf{L}\mathbf{c}|) \in \mathbb{R}^{2N-N_0},\ \mathbf{c} \in S_h\}$. Hence, S_h^{LP} is a non-empty compact set as the continuous image of the non-empty compact set S_h through ϕ. Moreover, S_h^{LP} is convex as the solution set of a linear program.

Next, (31) in Lemma 1 implies that $\phi : S_h \to S_h^{\mathrm{LP}}$ is a linear map. Moreover, ϕ is invertible and its inverse $\phi^{-1} : S_h^{\mathrm{LP}} \to S_h$ is also linear. The desired result immediately follows. □

F. Proof of Lemma 2

We first recap some useful properties of $\mathcal{M}_{\mathrm{L}}(\mathbb{R})$ given in [15]. Let $(\boldsymbol{\phi}, \mathbf{p})$ be a biorthogonal system for $\mathcal{N}_{\mathrm{L}}$ in the sense of Definition 3 in [15]. Therefore, $\mathbf{p} = (p_1, \dots, p_{N_0})$ is a basis of $\mathcal{N}_{\mathrm{L}}$ and $\boldsymbol{\phi} = (\phi_1, \dots, \phi_{N_0})$ is a basis of $\mathcal{N}_{\mathrm{L}}'$ the dual space of $\mathcal{N}_{\mathrm{L}}$. Referring to Part 2 of Theorem 5 in [15], any element $f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R})$ has a unique representation as $f = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{w\} + q$ where $w \in \mathcal{M}(\mathbb{R})$ and $q \in \mathcal{N}_{\mathrm{L}}$. The operator $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ is specified by Theorem 4 in [15] and is a right inverse of L such that $\boldsymbol{\phi}(\mathrm{L}_{\boldsymbol{\phi}}^{-1}\{w\}) = \mathbf{0}$ for any $w \in \mathcal{M}(\mathbb{R})$. Next, by Theorem 6 in [15], the predual of $\mathcal{M}_{\mathrm{L}}(\mathbb{R})$ is $C_{\mathrm{L}}(\mathbb{R}) = C_{\mathrm{L},\mathbf{p}}(\mathbb{R}) \oplus \mathcal{N}_{\mathrm{L}}'$, where $C_{\mathrm{L},\mathbf{p}}(\mathbb{R}) = \mathrm{L}^*\{C_0(\mathbb{R})\}$, L^* being the adjoint operator of L and $C_0(\mathbb{R})$ the set of continuous functions which vanish at infinity. Finally, we remind our reader that given a space $\mathcal{A}$ with predual $\mathcal{B}$ (*i.e.*, $\mathcal{B}' = \mathcal{A}$), the sequence $(a_k)_{k\in\mathbb{N}}$ where $a_k \in \mathcal{A}$ converges towards $a \in \mathcal{A}$ for the weak* topology if $\forall b \in \mathcal{B}$, $\lim_{k\to\infty}\langle a_k, b\rangle = \langle a, b\rangle$.

By Theorem 1, there exists a solution s to Problem (10) such that $\mathrm{L}\{s\} = \sum_{k=1}^K a_k\delta(\cdot - x_k)$ where $K \leq M - N_0$ and all x_k are pairwise distinct. As stated earlier, s can be represented as $s = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{w\} + p$ where $w = \sum_{k=1}^K a_k\delta(\cdot - x_k)$ and $p \in \mathcal{N}_{\mathrm{L}}$. We thus have $\mathcal{J}(s) = \mathcal{J}_0 = \|\boldsymbol{\nu}(s) - \mathbf{y}\|_2^2 + \lambda\|\mathbf{a}\|_1$. For a given $h > 0$, let $x_k^h \in h\mathbb{Z}$ be the grid element closest to x_k for all $k \in \{1,\dots,K\}$, *i.e.*, $|x_k - x_k^h| \leq \frac{h}{2}$. For small enough values of h, all x_k^h are pairwise distinct; we place ourselves in this configuration. We then define $s_h = p + \sum_{k=1}^K a_k \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{\delta(\cdot - x_k^h)\}$, which is in $\mathcal{M}_{\mathrm{L},h}(\mathbb{R})$ since it can also be written $s_h = q + \sum_{k=1}^K a_k\rho_{\mathrm{L}}(\cdot - x_k^h)$ where $q \in \mathcal{N}_{\mathrm{L}}$. It yields a cost $\mathcal{J}(s_h) = \|\boldsymbol{\nu}(s_h) - \mathbf{y}\|_2^2 + \lambda\|\mathbf{a}\|_1$ since the x_k^h are pairwise distinct.

Hence, there only remains to prove that $\boldsymbol{\nu}(s_h)$ converges to $\boldsymbol{\nu}(s)$ when $h \to 0$. We now show that $s_h \to s$ for the weak* topology when $h \to 0$; *i.e.*, $\langle s_h, f\rangle \to \langle s, f\rangle$ for any $f \in C_{\mathrm{L}}(\mathbb{R})$. Let $f = f_1 + f_2 \in C_{\mathrm{L}}(\mathbb{R})$ be the unique representation of f such that $f_1 \in C_{\mathrm{L},\mathbf{p}}(\mathbb{R})$ and $f_2 \in \mathcal{N}_{\mathrm{L}}'$. We first notice that $\boldsymbol{\phi}(s - s_h) = \mathbf{0}$ since $\boldsymbol{\phi}(\mathrm{L}_{\boldsymbol{\phi}}^{-1}\{w\}) = \mathbf{0}$ for any $w \in \mathcal{M}(\mathbb{R})$. Therefore, since $f_2 \in \mathrm{span}\{\phi_n\}_{n=1}^{N_0}$ we have $\langle s - s_h, f_2\rangle = 0$. Next, by definition of $C_{\mathrm{L},\mathbf{p}}(\mathbb{R})$, $\exists g \in C_0(\mathbb{R})$ such that $f_1 = \mathrm{L}^*\{g\}$. We thus have

$$\begin{aligned} \langle s - s_h, f_1\rangle &= \langle s - s_h, \mathrm{L}^*\{g\}\rangle \\ &= \langle \mathrm{L}\{s - s_h\}, g\rangle \\ &= \sum_{k=1}^K a_k\langle \delta(\cdot - x_k) - \delta(\cdot - x_k^h), g\rangle \\ &= \sum_{k=1}^K a_k(g(x_k) - g(x_k^h)). \end{aligned}$$

Moreover, from the definition of $C_0(\mathbb{R})$, g is continuous, and since $\lim_{h\to 0} x_k^h = x_k$, we have $\lim_{h\to 0}\langle s - s_h, f_1\rangle = 0$. We have thus proved that s_h converges to s for the weak* topology. Since $\boldsymbol{\nu}$ is weak*-continuous, we have $\lim_{h\to 0}\boldsymbol{\nu}(s_h) = \boldsymbol{\nu}(s)$ and thus $\lim_{h\to 0}\mathcal{J}(s_h) = \mathcal{J}(s)$.

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*. Cambridge University Press, 2012.

[3] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*. Birkhäuser Basel, 2013, vol. 1, no. 3.

[4] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[5] M. Unser, J. Fageot, and H. Gupta, "Representer theorems for sparsity-promoting ℓ_1 regularization," *IEEE Trans. Information Theory*, vol. 62, pp. 5167–5180, 2016.

[6] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[7] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589–592, 2008.

[8] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[9] 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 Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[10] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic resonance in medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[11] K. Bredies and H. Pikkarainen, "Inverse problems in spaces of measures," *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 19, no. 01, pp. 190–218, 2013.

[12] E. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[13] B. Adcock and A. Hansen, "Generalized sampling and infinite-dimensional compressed sensing," *Foundations of Computational Mathematics*, pp. 1–61, 2015.

[14] A. Flinth and P. Weiss, "Exact solutions of infinite dimensional total-variation regularized problems," *arXiv preprint arXiv:1708.02157*, 2017.

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

[16] S. Fisher and J. Jerome, "Spline solutions to L1 extremal problems in one and several variables," *Journal of Approximation Theory*, vol. 13, no. 1, pp. 73–83, 1975.

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

[18] I. Schoenberg, *Cardinal Spline Interpolation*. Philadelphia, PA: SIAM, 1973.

[19] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, vol. 16, no. 6, pp. 22–38, 1999.

[20] L. Schumaker, *Spline functions: basic theory*. Cambridge University Press, 2007.

[21] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[22] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5987–5999, 2013.

[23] B. Adcock, A. Hansen, C. Poon, and B. Roman, "Breaking the coherence barrier: A new theory for compressed sensing," in *Forum of Mathematics, Sigma*, vol. 5. Cambridge University Press, 2017.

[24] D. L. Donoho, "Superresolution via sparsity constraints," *SIAM journal on mathematical analysis*, vol. 23, no. 5, pp. 1309–1331, 1992.

[25] Q. Denoyelle, V. Duval, and G. Peyré, "Support recovery for sparse super-resolution of positive measures," *Journal of Fourier Analysis and Applications*, vol. 23, no. 5, pp. 1153–1194, 2017.

[26] V. Duval and G. Peyré, "Exact support recovery for sparse spikes deconvolution," *Foundations of Computational Mathematics*, vol. 15, no. 5, pp. 1315–1355, 2015.

[27] ——, "Sparse spikes deconvolution on thin grids," *arXiv preprint arXiv:1503.08577*, 2015.

[28] N. Boyd, T. Hastie, S. Boyd, B. Recht, and M. Jordan, "Saturating splines and feature selection," *arXiv preprint arXiv:1609.06764*, 2016.

[29] C. Dossal, V. Duval, and C. Poon, "Sampling the Fourier transform along radial lines," *SIAM Journal on Numerical Analysis*, vol. 55, no. 6, pp. 2540–2564, 2017.

[30] C. Aubel, D. Stotz, and H. Bölcskei, "A theory of super-resolution from short-time fourier transform measurements," *Journal of Fourier Analysis and Applications*, vol. 24, no. 1, pp. 45–107, 2018.

[31] M. Unser and T. Blu, "Cardinal exponential splines: Part I-theory and filtering algorithms," *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1425–1438, 2005.

[32] L. Schwartz and I. de mathématique (Strasbourg), *Théorie des distributions*. Hermann Paris, 1957, vol. 2.

[33] M. Unser and P. D. Tafti, *An introduction to sparse stochastic processes*. Cambridge University Press, 2014.

[34] W. Rudin, *Real and complex analysis*. Tata McGraw-Hill Education, 1987.

[35] M. Arigovindan, M. Suhling, P. Hunziker, and M. Unser, "Variational image reconstruction from arbitrarily spaced samples: A fast multiresolution spline solution," *IEEE Transactions on Image Processing*, vol. 14, no. 4, pp. 450–460, 2005.

[36] R. J. Tibshirani *et al.*, "The LASSO problem and uniqueness," *Electronic Journal of Statistics*, vol. 7, pp. 1456–1490, 2013.

[37] G. Dantzig, A. Orden, and P. Wolfe, "The generalized simplex method for minimizing a linear form under linear inequality restraints," *Pacific Journal of Mathematics*, vol. 5, no. 2, pp. 183–195, 1955.

[38] D. G. Luenberger, *Introduction to linear and nonlinear programming*. Addison-Wesley Reading, MA, 1973, vol. 28.

[39] M. Unser, E. Soubies, F. Soulez, M. McCann, and L. Donati, "GlobalBioIm: A unifying computational framework for solving inverse problems," in *Proceedings of the OSA Imaging and Applied Optics Congress on Computational Optical Sensing and Imaging (COSI'17)*, San Francisco CA, USA, June 26-29, 2017, paper no. CTu1B.

[40] L. Gurobi Optimization, "Gurobi optimizer reference manual," 2018. [Online]. Available: http://www.gurobi.com

[41] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE transactions on information theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[42] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock, "An introduction to total variation for image analysis," *Theoretical foundations and numerical methods for sparse recovery*, vol. 9, no. 263-340, p. 227, 2010.