Journal of Computational and Applied Mathematics 406 (2022) 114044

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Sparsest piecewise-linear regression of one-dimensional data

Thomas Debarre [a,*], Quentin Denoyelle [a], Michael Unser [a], Julien Fageot [b]

[a] EPFL, Biomedical Imaging Group, Lausanne, Switzerland
[b] EPFL, AudioVisual Communications Laboratory, Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 12 August 2021
Received in revised form 25 November 2021

Keywords:
Inverse problems
Total-variation norm for measures
Sparsity
Splines

ABSTRACT

We study the problem of one-dimensional regression of data points with total-variation (TV) regularization (in the sense of measures) on the second derivative, which is known to promote piecewise-linear solutions with few knots. While there are efficient algorithms for determining such adaptive splines, the difficulty with TV regularization is that the solution is generally non-unique, an aspect that is often ignored in practice. In this paper, we present a systematic analysis that results in a complete description of the solution set with a clear distinction between the cases where the solution is unique and those, much more frequent, where it is not. For the latter scenario, we identify the sparsest solutions, i.e., those with the minimum number of knots, and we derive a formula to compute the minimum number of knots based solely on the data points. To achieve this, we first consider the problem of exact interpolation which leads to an easier theoretical analysis. Next, we relax the exact interpolation requirement to a regression setting, and we consider a penalized optimization problem with a strictly convex data-fidelity cost function. We show that the underlying penalized problem can be reformulated as a constrained problem, and thus that all our previous results still apply. Based on our theoretical analysis, we propose a simple and fast two-step algorithm, agnostic to uniqueness, to reach a sparsest solution of this penalized problem.
© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Regression problems consist in learning a function f that best approximates some data $(x_m, y_m)_{m=1}^M$, where M is the number of data points, in the sense that $f(x_m) \approx y_m$. This is typically achieved by parametrizing f with a vector of parameters $\boldsymbol{\theta}$, and minimizing some objective function with respect to $\boldsymbol{\theta}$. The oldest and most basic form or regression is linear regression: f is parametrized as a linear (or affine) function. Although this model has the advantage of being very simple, it is very limited due to the fact that many data distributions are poorly approximated by linear functions, as illustrated by the dotted line example in Fig. 1. The choice of parametrization $\boldsymbol{\theta}$ is therefore crucial, as it must strike an appropriate balance between two conflicting desirable properties. Firstly, in order to be suitable for a variety of problems, the parametric model should be flexible enough to represent a large class of functions. In the field of machine learning, where regression is known as *supervised learning*, this quest for universality is for instance highlighted by several universal approximation theorems for artificial neural networks [1–3]. Next, the model should be simple enough so that it generalizes well to input vectors $\mathbf{x}$ that are outside of the training set. Indeed, a known pitfall of machine learning algorithms is overfitting, which happens when the model is unduly complex and fits too closely to the training data

* Corresponding author.
E-mail addresses: thomas.debarre@gmail.com (T. Debarre), quentin.denoyelle@epfl.ch (Q. Denoyelle), michael.unser@epfl.ch (M. Unser), julien.fageot@epfl.ch (J. Fageot).

https://doi.org/10.1016/j.cam.2021.114044
0377-0427/© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

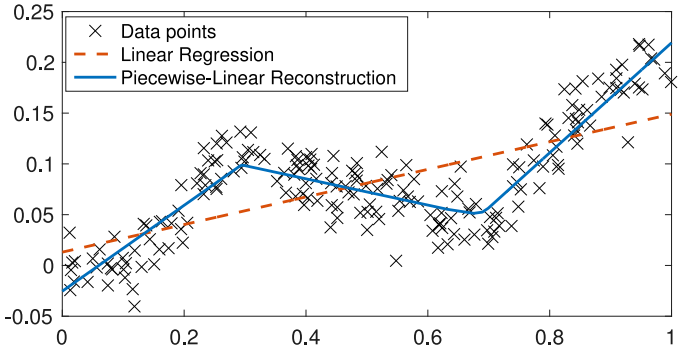


Fig. 1. Examples of reconstructions.

[4, Chapter 3]. This leads to poor generalization abilities for out-of-sample data. This pitfall is often dealt with by adding some regularization to the objective function, which tends to simplify the model. The overarching guiding principle to avoid overfitting is Occam's razor: the simplest model that explains the data well will generalize better and should thus be selected.

1.1. Problem formulation

In this paper, we study the regression (or supervised learning) problem in one dimension, *i.e.*, $f : \mathbb{R} \to \mathbb{R}$ and $x_m, y_m \in \mathbb{R}$. However, instead of parametrizing the reconstructed function, we formulate the learning problem as a regularized inverse problem in a continuous-domain framework. Inspired by their connection (that we discuss later on) to popular ReLU (rectified linear unit) neural networks, we focus on reconstructing piecewise-linear splines. Our metric for model simplicity is sparsity, *i.e.*, the number of spline knots. For regularization purposes, we therefore use the total variation (TV) norm for measures $\|\cdot\|_{\mathcal{M}}$, which is defined over the space of bounded Radon measures $\mathcal{M}(\mathbb{R})$. This norm is known to promote sparse solutions in the desired sense, as will be clarified in (1). We formulate the following optimization problem, which we refer to as the *generalized Beurling LASSO* (g-BLASSO)

$$\arg\min_{f} \sum_{m=1}^{M} E(f(x_m), y_m) + \lambda \|\mathrm{D}^2 f\|_{\mathcal{M}}, \tag{g-BLASSO}$$

where E is a cost function that penalizes the fidelity of $f(x_m)$ to the data $y_m \in \mathbb{R}$ (*e.g.*, a quadratic loss $E(z, y) = \frac{1}{2}(z-y)^2$). We assume that the sampling locations are ordered, *i.e.*, $x_1 < \cdots < x_M$. The parameter $\lambda > 0$ balances the contribution of the data fidelity and the regularization, and D^2 is the second-derivative operator. The terminology generalized Beurling LASSO comes from the Beurling LASSO (BLASSO) which is used in the Dirac recovery literature [5]. Indeed, the (g-BLASSO) problem is a generalization of the BLASSO due to the presence of a regularization operator D^2, which is not present in the latter problem. It is known [6–8] that the extreme points solutions to the (g-BLASSO) are piecewise-linear splines of the form

$$f_{\mathrm{opt}}(x) = b_0 + b_1 x + \sum_{k=1}^{K} a_k (x - \tau_k)_+, \tag{1}$$

where $x_+ = \max(0, x)$ is the ReLU, $b_0, b_1, a_k, \tau_k \in \mathbb{R}$, and the number of spline knots K is bounded by $K \leq M - 2$. This representer theorem has two important components:

- the (g-BLASSO) has solutions of the prescribed form, *i.e.*, piecewise-linear splines. This stems from the choice of the regularization, *i.e.*, the TV norm of the second derivative;
- the sparsity is bounded by the number of training data by $K \leq M - 2$.

In terms of model simplicity, the bound $K \leq (M - 2)$ is typically uninformative in machine learning problems: in Fig. 1, it yields $K \leq M - 2 = 198$, which is clearly much higher than the desired sparsity. However, this bound does not take the effect of the regularization parameter λ into account. Indeed, $\lambda \to 0$ will roughly lead to a learned function f that interpolates all the data points, with typically close to $K = M - 2$ knots. At the other extreme, the limit $\lambda \to +\infty$ leads to linear regression and thus sparsity $K = 0$ due to the fact that linear functions are not penalized by the regularization. Therefore, the interesting case is the intermediate regime (as illustrated by the solid curve in Fig. 1), in which the overall trend is that the sparsity K decreases as λ increases. Hence, λ controls the universality *versus* simplicity trade-off.

1.2. Summary of contributions and outline

The above purely qualitative observation is far from telling the whole story. In particular, it does not prescribe how λ should be chosen in practice. We attempt to overcome this impediment by giving a full description of the solution set of the (g-BLASSO). The basis of our analysis is the classical observation (see for instance [7, Theorem 5]) that when E is strictly convex, there exists a unique vector $\mathbf{y}_\lambda = (y_{\lambda,1}, \dots, y_{\lambda,M}) \in \mathbb{R}^M$ such that the (g-BLASSO) is equivalent to the constrained problem

$$\underset{\substack{f: f(x_m)=y_{\lambda,m},\\ m\in\{1,\dots,M\}}}{\arg\min} \ \|\mathrm{D}^2 f\|_{\mathcal{M}}, \tag{g-BPC}$$

which we refer to as the *generalized basis pursuit in the continuum* (g-BPC).[1] Our terminology is inspired by the (discrete) basis pursuit (BP) [11], which is also a constrained problem; as for the (g-BLASSO), the "generalization" is due to the presence of a regularization operator D^2 which is absent in the BP. We therefore carry out our theoretical analysis on the more straightforward (g-BPC) problem, and we attest that these results apply to the (g-BLASSO) as well, provided that $\mathbf{y}_\lambda$ is known. For this analysis, we use mathematical tools based on duality theory, and we exploit the very specific form of the so-called dual certificate for our regularization operator D^2. We describe in a systematic way the form of the solution set and identify the set of sparsest solutions. The fact the optimization problems with sparsity-promoting regularization sometimes have multiple solutions is often sidestepped in the literature by identifying specific cases of uniqueness [12–14]. When it is not, existing works typically provide the form of certain solutions [6,15], but they do not characterize cases of uniqueness nor do they give a *complete* description of the solution set as we do here. Concerning our specific problem, it is known that the function that simply connects the points $(x_1, y_{0,1}), \dots, (x_M, y_{0,M})$ is always a solution to the (g-BPC) (see [16, Theorem 1] and [17, Proposition 7]). We refer to it as the *canonical solution*. Building on this result, our contributions on the theoretical and algorithmic sides concerning the (g-BLASSO) are summarized below.

1. **Theory**
   Our main theoretical contributions are the following.
   - In Section 3, we fully describe the solution set of the (g-BPC) by specifying the intervals in which all solutions follow the canonical solution, and those in which they do not (Theorem 2). This allows us to characterize the cases where the (g-BPC) admits a unique solution. When they differ, we give a geometrical description of the set in which the graph of all solutions lies in Theorem 3.
   - When there are multiple solutions, the canonical solution can be made sparser in certain regions, which is the topic of Section 4. More precisely, in Theorem 4, we express the minimum achievable sparsity of a solution to the (g-BPC) as a simple function of $\mathbf{x} \overset{\text{def.}}{=} (x_1, \dots, x_M)$ and $\mathbf{y}_0$, which we denote by $K_{\min}(\mathbf{x}, \mathbf{y}_0)$. Concerning the solution set, we fully describe the set of sparsest solutions of the (g-BPC). In particular, we characterize the cases of uniqueness, and provide a description of the sparsest solutions together with the number of degrees of freedom $n_{\text{free}}(\mathbf{x}, \mathbf{y}_0)$, that we characterize and show to be finite.
   - In Section 5.1, we extend the results of the first two items to the (g-BLASSO). This is a consequence of the aforementioned equivalence between the (g-BLASSO) and the (g-BPC) problems, given in Proposition 7. We also specify the limit value $\lambda_{\max}$, for which any $\lambda \geq \lambda_{\max}$ amounts to linear regression in Proposition 10.
2. **Algorithm**
   These theoretical findings warrant our simple and fast algorithm, presented in Section 5.2, for reaching (one of) the sparsest solution(s) to the (g-BLASSO). The algorithm, which is agnostic to uniqueness, is divided in two parts: first, we compute the $\mathbf{y}_\lambda$ vector for the (g-BPC) problem by solving a standard discrete ℓ_1-regularized problem. Next, we find a sparsest solution to the (g-BLASSO) (with sparsity $K_{\min}(\mathbf{x}, \mathbf{y}_\lambda)$) by optimally sparsifying its canonical solution in some prescribed regions that are determined by our theoretical results. This sparsification step is detailed in Algorithm 1 and has complexity $\mathcal{O}(M)$.
   This complete algorithm provides a simple and fast way for the user to judiciously choose λ by evaluating the data fidelity loss $\sum_{m=1}^{M} E(f(x_m), y_m)$ *versus* the optimal sparsity $K_{\min}(\mathbf{x}, \mathbf{y}_0)$—which depends on λ—as a proxy for the universality *versus* simplicity trade-off. We illustrate this in our experiments in Section 6. The value of λ may vary between $\lambda \to 0$ (which at the limit amounts to the (g-BPC) problem) and an upper bound $\lambda = \lambda_{\max}$ mentioned above. Note that existing algorithms that solve the (g-BLASSO) such as that introduced in [17] are a lot more complex and computationally expensive. Moreover, to the best of our knowledge, no existing algorithm has the guarantee of reaching a sparsest solution of the (g-BPC) or the (g-BLASSO).

[1] A similar terminology, the "continuous basis pursuit", is used in a different context in [9,10].

1.3. Related works

Discrete ℓ_1 optimization. Putting aside for now the regularization operator D^2, the optimization problems (g-BPC) and (g-BLASSO) are the continuous-domain counterparts of the basis pursuit [11] and the LASSO [18], which were introduced in the late 90s. These problems are the precursors of the type of ℓ_1-recovery techniques used in compressed sensing [19–23]. These approaches provide solutions with only few nonzero coefficients. They are at the cornerstone of sparse statistical learning [24] and sparse signal processing [25]. Theoretical recovery guarantees have been proved, see for example [26]; however it is worth noting that in their initial formulations, these methods are inherently discrete and therefore adapted to recover finite-dimensional physical quantities.

Reconstruction in infinite-dimensional spaces. In our context, we aim at learning a continuous-domain function $f : \mathbb{R} \to \mathbb{R}$ from finite-dimensional data (the values $y_m = f(x_m)$ for $m \in \{1, \dots, M\}$). It is therefore natural to formulate the optimization task in infinite dimension to perform the reconstruction. The problem is then inherently ill-posed: not only is the system undetermined, as it is also the case in compressed sensing, but we have infinitely many degrees of freedom with finitely many constraints for the reconstruction. Kernel methods based on quadratic regularization are an elegant way of removing this ill-posedness [27], with the effect of restricting the approximation to a finite-dimensional subset of a Hilbert space [28–30]. The challenge is then to choose this Hilbert space adequately. These approaches are fruitful, but they still ultimately revert to the finite-dimensional setting. Taking inspiration from ℓ_1-based methods for sparse vectors, new approaches have been proposed that go beyond the Hilbert space setting, such as [31–34].

Reconstruction in measure spaces. A fertile continuous-domain problem to which discrete ℓ_1 methods were recently adapted is sparse spikes deconvolution [5,12,13,35,36]. The aim is to recover sums of Dirac masses (point sources signals) over a continuous domain by extending the ℓ_1 regularization to a gridless setup thanks to the total variation norm $\|\cdot\|_{\mathcal{M}}$, which is defined over the space of Radon measures $\mathcal{M}(\mathbb{R})$. The underlying optimization problems, either formulated in a constrained form in the noiseless case [12] or in a penalized form known as the BLASSO [5] in the presence of noise, are thus solved over a nonreflexive Banach space. The role of the total variation norm in variational methods has a rich history [37,38] (see [8, Section 1] for additional references). From a theoretical standpoint, many reconstruction guarantees are proved, such as exact recovery of discrete measures (sums of Dirac masses) in the noiseless case [12,14], robustness to noise [36,39–41], support recovery [10,13,42,43] and super-resolution for positive discrete measures [5,44–49].

From a numerical standpoint, there exist several different strategies to solve these problems. A first one is based on spatial discretization which leads back to the LASSO and algorithms such as FISTA [50]. There are also greedy algorithms such as continuous-domain Orthogonal Matching Pursuit (OMP) [51]. In special setups (typically Fourier measurements), it is possible to reformulate the optimization problems as semidefinite programs [12,52,53]. Finally, recent developments based on the Frank–Wolfe (FW) algorithm [54] solve the BLASSO directly over the space of Radon measures [36]. These FW-based methods improve on the traditional FW algorithm due to the possibility of moving the spikes in the continuous domain to further decrease the objective function [55–58].

From Dirac masses recovery to spline reconstruction. More generally, Dirac masses recovery is part of a trend that promotes continuous-domain formalisms for signal reconstruction. By adding a differential operator to the total variation regularization, one allows for more diverse reconstructions than the recovery of sums of Dirac impulses, while keeping the sparsity-promoting effect of the total variation norm. Even predating the era of ReLU networks, the (g-BLASSO) and, to a greater degree, the (g-BPC)—or variations thereof—have been of keen interest to the signal processing and statistics communities. Adding a differential operator leads to spline reconstructions, a result that can be traced back to [15,59] in the 70s. In [60], Pinkus proved that the canonical solution—that simply connects the data points—is the unique solution to the (g-BPC) in some special cases, a result that we recover in our analysis. Later, Koenker et al. [16, Theorem 1] and Mammen and Van de Geer [17, Proposition 7] proved that the canonical solution is indeed a solution to the (g-BPC). These works also propose algorithms to solve the (g-BLASSO) for any value of λ. However, contrary to this paper, none of the aforementioned works describe the full solution set of the (g-BPC), nor identifies its sparsest solutions. There has been a promising new surge of very recent works on related problems, both on the theoretical and the algorithmic sides [8,61–67]. Several very general theories, that incorporate the (g-BLASSO) and the (g-BPC), and that deal with optimization in Banach spaces with various differential regularization operators, have also been recently developed [8,64,68,69].

Relu networks, piecewise-linear splines, and the (g-BLASSO). A modern approach to supervised learning is neural networks, which in recent years have become the gold standard for an impressive number of applications [70]. Many recent papers have highlighted the property that today's state-of-the-art convolutional neural networks (CNNs) with rectified linear unit (ReLU) activations specify an input–output relation $f : \mathbb{R}^d \to \mathbb{R}$, where d is the number of dimensions, that is continuous and piecewise-linear (CPWL) [71–73]. This result stems from the fact that the ReLU nonlinearity is itself a CPWL function, as well as, for instance, the widespread max-pooling operation. In fact, there are indications that using more general piecewise-linear splines as activation functions could be more effective than restricting to the ReLU or leaky ReLU [61,74,75]. In the one-dimensional case $d = 1$, it follows that the learned function of a ReLU network is a piecewise-linear spline [76], just like the solutions to the (g-BLASSO) given by (1). The trade-off between universality and Occam's razor is then determined by the network size and architecture. Many recent papers in the literature have investigated this

connection between ReLU networks and piecewise-linear splines [77,78], including universality properties [76,79,80]. We also mention [81], which considers more general spline activation functions.

Moreover, several works have specifically underscored the relevance of the (g-BLASSO)—or related problems [82]—in machine learning by showing that it is equivalent to the training of a one-dimensional ReLU network with standard weight decay [83,84]. Therefore, although the current trend of overparametrizing neural networks is somewhat antagonistic to our paradigm of sparsity, our full description of the solution set of the (g-BLASSO) (including its non-sparse solutions) could be relevant to the neural network community. Others recent works have designed multidimensional ($d > 1$) equivalents of the regularization term $\|\mathrm{D}^2 f\|_{\mathcal{M}}$ and derive similar connections to neural networks [85,86].

2. Mathematical preliminaries

The task of recovering a continuous-domain function from finitely many samples is obviously ill-posed; this issue is commonly addressed by adding a regularization term. As a regularization norm, we consider $\|\cdot\|_{\mathcal{M}}$, which is the continuous-domain counterpart of the ℓ_1-norm, and is known to promote sparse solutions [22]. Some of the results of this section (in Sections 2.2 and 2.3) are not new, as they can be seen as a special case of the general framework developed in previous works [6,7,69] to the case of the second-derivative operator $\mathrm{L} = \mathrm{D}^2$. Nevertheless, we provide a self-contained treatment, for the benefit of readers who are unfamiliar with the general theory.

2.1. The measure space $\mathcal{M}(\mathbb{R})$

We denote by $\mathcal{M}(\mathbb{R})$, the space of bounded Radon measures on $\mathbb{R}$. It is a nonreflexive Banach space and is defined as the topological dual of the space $\mathcal{C}_0(\mathbb{R})$ of continuous functions that vanish at $\pm\infty$ endowed with the supremum norm $\|\cdot\|_\infty$. The duality product between a measure $w \in \mathcal{M}(\mathbb{R})$ and a function $f \in \mathcal{C}_0(\mathbb{R})$ is denoted by $\langle w, f\rangle \overset{\text{def.}}{=} \int_{\mathbb{R}} f \mathrm{d}w$. The norm on $\mathcal{M}(\mathbb{R})$ is called the total-variation norm and is given by

$$\forall w \in \mathcal{M}(\mathbb{R}), \quad \|w\|_{\mathcal{M}} \overset{\text{def.}}{=} \sup_{f \in \mathcal{C}_0(\mathbb{R}),\ \|f\|_\infty \leq 1} \langle w, f\rangle. \tag{2}$$

Moreover, we have the continuous embeddings

$$\mathcal{S}(\mathbb{R}) \subseteq \mathcal{M}(\mathbb{R}) \subseteq \mathcal{S}'(\mathbb{R}), \tag{3}$$

where $\mathcal{S}(\mathbb{R})$ is the Schwartz space of smooth and rapidly decaying functions and $\mathcal{S}'(\mathbb{R})$ is its topological dual, the space of tempered distributions [87]. We observe that we can replace $\mathcal{C}_0(\mathbb{R})$ by $\mathcal{S}(\mathbb{R})$ in (2), by invoking the denseness of $\mathcal{S}(\mathbb{R})$ in $\mathcal{C}_0(\mathbb{R})$, and then characterize the bounded Radon measures among $\mathcal{S}'(\mathbb{R})$ as

$$\mathcal{M}(\mathbb{R}) = \{w \in \mathcal{S}'(\mathbb{R}) : \sup_{f \in \mathcal{S}(\mathbb{R}),\ \|f\|_\infty \leq 1} \langle w, f\rangle < \infty\}. \tag{4}$$

2.2. The native space $\mathrm{BV}^{(2)}(\mathbb{R})$

Motivated by the form of the regularization in the (g-BPC) and the (g-BLASSO), we introduce the space over which we shall optimize both problems. It is defined as

$$\mathrm{BV}^{(2)}(\mathbb{R}) \overset{\text{def.}}{=} \{f \in \mathcal{S}'(\mathbb{R}) : \mathrm{D}^2 f \in \mathcal{M}(\mathbb{R})\}, \tag{5}$$

with $\mathrm{D}^2 : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$ the second-derivative operator. The space $\mathrm{BV}^{(2)}(\mathbb{R})$ has been considered and studied in [61, Section 2.2]. It is the second-order generalization of the well-known space of functions with bounded variation. For the sake of completeness, a detailed presentation of the mathematical properties of $\mathrm{BV}^{(2)}(\mathbb{R})$ is provided in Appendix A. For now, it is important to remember that $\mathrm{BV}^{(2)}(\mathbb{R})$ is a Banach space equipped with the norm

$$\|f\|_{\mathrm{BV}^{(2)}} \overset{\text{def.}}{=} \|\mathrm{D}^2 f\|_{\mathcal{M}(\mathbb{R})} + \sqrt{f(0)^2 + (f(1) - f(0))^2}. \tag{6}$$

Moreover, any function $f \in \mathrm{BV}^{(2)}(\mathbb{R})$ is continuous and such that $f(x) = \mathcal{O}(x)$ at infinity (see Proposition 11 in Appendix A).

For any $w \in \mathcal{M}(\mathbb{R})$, we denote by $\mathrm{D}_0^{-2}\{w\}$ the unique function $f \in \mathrm{BV}^{(2)}(\mathbb{R})$ such that $\mathrm{D}^2 f = w$ and $f(0) = f(1) = 0$, according to the last point of Proposition 11. Then, D_0^{-2} is a continuous operator from $\mathcal{M}(\mathbb{R})$ to $\mathrm{BV}^{(2)}(\mathbb{R})$, whose main properties are summarized in Proposition 12 in Appendix A. Its effect is to doubly integrate the measure on which it operates.[2] Moreover, any $f \in \mathrm{BV}^{(2)}(\mathbb{R})$ can be uniquely decomposed as

$$\forall x \in \mathbb{R}, \quad f(x) = \mathrm{D}_0^{-2}\{w\}(x) + \beta_0 + \beta_1 x, \tag{7}$$

[2] The notation D_0^{-2} has two justifications. First, it recalls that this operator is a right-inverse of the second derivative D^2. However, the index 0 indicates that D_0^{-2} is *not* a left-inverse, as revealed by Proposition 12 in Appendix A.

where $w \in \mathcal{M}(\mathbb{R})$ and $\beta_0, \beta_1 \in \mathbb{R}$ satisfy

$$w = \mathrm{D}^2 f, \quad \beta_0 = f(0), \quad \text{and} \quad \beta_1 = f(1) - f(0). \tag{8}$$

We call the measure w the *innovation* of f. The key elements of $\mathrm{BV}^{(2)}(\mathbb{R})$ we are interested in are piecewise-linear splines, which are defined as follows.

Definition 1 (*Piecewise-Linear Spline*)**.** A piecewise-linear spline is a function $f \in \mathrm{BV}^{(2)}(\mathbb{R})$ whose innovation $w = \mathrm{D}^2 f \in \mathcal{M}(\mathbb{R})$ is a weighted sum of Dirac masses $w = \sum_{k=1}^{K} a_k \delta(\cdot - \tau_k)$, where $K \in \mathbb{N}$ is the number of knots (*i.e.*, singularities), called the *sparsity* of the spline and $a_k, \tau_k \in \mathbb{R}$.

It follows from Definition 1 that a piecewise-linear spline f can equivalently be written as

$$f(x) = b_0 + b_1 x + \sum_{k=1}^{K} a_k (x - \tau_k)_+, \tag{9}$$

where $b_0, b_1 \in \mathbb{R}$. Note that this representation is different from that of (7) (in general, $(\beta_0, \beta_1) \neq (b_0, b_1)$); however we favor the representation (9) for splines due to its simplicity.

2.3. Representer theorem for $\mathrm{BV}^{(2)}(\mathbb{R})$

The native space $\mathrm{BV}^{(2)}(\mathbb{R})$ allows us to precisely define the optimization problems we are interested in. Indeed, it is the largest space for which the regularization $\|\mathrm{D}^2 f\|_{\mathcal{M}}$ is well-defined and finite. The following result is a special case of a more general theory, which is now well established.

Theorem 1 (*Representer Theorem for* $\mathrm{BV}^{(2)}(\mathbb{R})$)**.** *Let* $\mathbf{x} = (x_1, \ldots, x_M) \in \mathbb{R}^M$ *be a collection of distinct* $M \geq 2$ *ordered sampling locations and* $\mathbf{y}_0 \in \mathbb{R}^M$*. We consider the set of solutions*

$$\mathcal{V}_0 \overset{\text{def.}}{=} \operatorname*{arg\,min}_{\substack{f \in \mathrm{BV}^{(2)}(\mathbb{R}) \\ f(x_m) = y_{0,m},\ m=1,\ldots,M}} \|\mathrm{D}^2 f\|_{\mathcal{M}}. \tag{g-BPC}$$

Moreover, we fix $\lambda > 0$ *and* $\mathbf{y} \in \mathbb{R}^M$*, together with a cost function* $E : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$ *such that* $E(\cdot, y)$ *is strictly convex, coercive, and differentiable for any* $y \in \mathbb{R}$ *and* $\lambda > 0$*. We also consider the set of solutions*

$$\mathcal{V}_\lambda \overset{\text{def.}}{=} \operatorname*{arg\,min}_{f \in \mathrm{BV}^{(2)}(\mathbb{R})} \sum_{m=1}^{M} E(f(x_m), y_m) + \lambda \|\mathrm{D}^2 f\|_{\mathcal{M}}. \tag{g-BLASSO}$$

Then, for any $\lambda \geq 0$ *(including 0),* $\mathcal{V}_\lambda$ *is nonempty, convex, and weak-* compact in* $\mathrm{BV}^{(2)}(\mathbb{R})$*, and is the weak-* closure of the convex hull of its extreme points. The latter are all piecewise-linear splines of the form*

$$f_{\text{extreme}}(x) = b_0 + b_1 x + \sum_{k=1}^{K} a_k (x - \tau_k)_+, \tag{10}$$

where $b_0, b_1 \in \mathbb{R}$*, the weights* a_k *are nonzero, the knots locations* $\tau_k \in \mathbb{R}$ *are distinct, and* $K \leq M - 2$*.*

Following the seminal work of Fisher and Jerome [15], this result was proved in [6, Theorem 2] for $\lambda = 0$ and for a general spline-admissible operator L in the regularization term $\|\mathrm{L} \cdot\|_{\mathcal{M}(\mathbb{R})}$. The case $\lambda > 0$ is proved in [7, Theorem 4] for a general cost function E, by reducing the analysis to the optimization problem (g-BPC) (as we shall do in Section 5). Theorem 1 is then a particular case of these two works for the regularization operator $\mathrm{L} = \mathrm{D}^2$, whose null space is generated by $x \mapsto 1$ and $x \mapsto x$, and for sampling measurements. Note that the application of these known theorems requires to prove that the point evaluation $f \mapsto f(x_0)$ is weak-* continuous on $\mathrm{BV}^{(2)}(\mathbb{R})$ for any $x_0 \in \mathbb{R}$, which has been shown in [61, Theorem 1]. These theorems have been recently revisited and/or extended by several authors [8,63,64].

Theorem 1 is called a "representer theorem", as initially proposed in [6], because it specifies the form of the extreme-point solutions of the optimization problem. It is then possible to reduce the optimization task to functions of the form (10), which considerably simplifies the analysis [7,88]. Theorem 1 is also an existence result. It guarantees that the minimization problem (g-BPC) admits at least one piecewise-linear solution. In particular, if the solution is unique, then it is a piecewise-linear spline. However, Theorem 1 is not informative regarding the knots locations τ_k, which may be distinct from the sampling locations x_m.

To the best of our knowledge, very few attempts have been made to characterize the cases where gTV optimization problems admit a unique solution, and to describe the solution set when the solution is not unique. In this paper, we provide complete answers to these questions for the reconstruction of functions via sampling measurements and with $\mathrm{BV}^{(2)}$-type regularization.

2.4. Dual certificates

This section presents the main tools for the study of the (g-BPC) problem (with $\mathbf{x} \in \mathbb{R}^M$ the ordered distinct sampling locations and $\mathbf{y}_0 \in \mathbb{R}^M$ the measurements), coming from the duality theory, which are at the core of our contributions. Our strategy consists in studying a particular class of continuous functions, called *dual certificates*, which can be used individually to certify that an element $f \in \mathrm{BV}^{(2)}(\mathbb{R})$ is a solution of the optimization problem (g-BPC). More interestingly, from the properties of a given dual certificate, it is possible to precisely describe the whole structure of the set of solutions (see Theorem 2) and, in particular, to determine whether or not the sparse solution given by Theorem 1 is the unique solution of the problem (see Proposition 6).

Before giving the main results of this section (Propositions 1 and 2), let us first introduce the definition of a dual pre-certificate.

Definition 2 (*Dual Pre-Certificate*)**.** We say that a function $\eta \in \mathcal{C}_0(\mathbb{R})$ is a *dual pre-certificate* (for the problem (g-BPC)) if its norm satisfies $\|\eta\|_\infty \leq 1$ and if η is of the form

$$\eta = \sum_{m=1}^{M} c_m (x_m - \cdot)_+ \tag{11}$$

for some vector $\mathbf{c} = (c_1, \ldots, c_M) \in \mathbb{R}^M$ such that $\langle \mathbf{c},\ \mathbf{1} \rangle = \langle \mathbf{c},\ \mathbf{x} \rangle = 0$ (with $\mathbf{1} \overset{\text{def.}}{=} (1, \ldots, 1) \in \mathbb{R}^M$).

A dual pre-certificate is therefore a piecewise-linear spline. The conditions $\langle \mathbf{c},\ \mathbf{1} \rangle = \langle \mathbf{c},\ \mathbf{x} \rangle = 0$ ensure that η is compactly supported, and is thus an element of $\mathcal{C}_0(\mathbb{R})$ (indeed, we have $\eta(x) = -\langle \mathbf{c}, \mathbf{1} \rangle x + \langle \mathbf{c}, \mathbf{x} \rangle = 0$ for any $x \leq x_1$). We shall present an explicit construction of such a pre-certificate in Proposition 4 with the piecewise-linear spline η_{cano}. A dual certificate is a pre-certificate that satisfies an additional condition (see Proposition 1) that ensures that the vector $\mathbf{c} \in \mathbb{R}^M$ in Definition 2 is a solution of the dual problem of (g-BPC).

From (7), we know we can parametrize any $f \in \mathrm{BV}^{(2)}(\mathbb{R})$ with a unique element $(w, (\beta_0, \beta_1)) \in \mathcal{M}(\mathbb{R}) \times \mathbb{R}^2$ through the relation

$$\forall x \in \mathbb{R}, \quad f(x) = \mathrm{D}_0^{-2}\{w\}(x) + \beta_0 + \beta_1 x. \tag{12}$$

Dual certificates determine the localization of the support of w when f is a solution of (g-BPC). To formulate this property, we need the following definition which introduces the concepts of signed support of a measure (see Section 1.4 of [13]) and signed saturation set of a pre-certificate (see [13, Definition 3]).

Definition 3 (*Signed Support and Signed Saturation Set*)**.** Let $w \in \mathcal{M}(\mathbb{R})$ and $\eta \in \mathcal{C}_0(\mathbb{R})$ be a dual pre-certificate in the sense of Definition 2. We define the *signed support* of w by

$$\mathrm{supp}_\pm(w) \overset{\text{def.}}{=} \mathrm{supp}(w_+) \times \{1\} \cup \mathrm{supp}(w_-) \times \{-1\}, \tag{13}$$

where w_+ and w_- are positive measures coming from the Jordan decomposition of $w = w_+ - w_-$. Moreover from the positive and negative saturation sets of η, defined as

$$\mathrm{sat}_+(\eta) \overset{\text{def.}}{=} \{x \in \mathbb{R} : \eta(x) = 1\} \quad \text{and} \quad \mathrm{sat}_-(\eta) \overset{\text{def.}}{=} \{x \in \mathbb{R} : \eta(x) = -1\} \tag{14}$$

respectively, we define the *signed saturation set* of η by

$$\mathrm{sat}_\pm(\eta) \overset{\text{def.}}{=} \mathrm{sat}_+(\eta) \times \{1\} \cup \mathrm{sat}_-(\eta) \times \{-1\}. \tag{15}$$

Note that the sets $\mathrm{supp}_\pm(w)$, $\mathrm{sat}_+(\eta)$, $\mathrm{sat}_-(\eta)$, $\mathrm{sat}_\pm(\eta)$ are all closed. A dual pre-certificate η is a piecewise-linear spline in $\mathcal{C}_0(\mathbb{R})$ with norm $\|\eta\|_\infty \leq 1$. Hence, its signed saturation set is necessarily a union of closed intervals (that can be singletons).

We can now state the first main result of this section, the proof of which can be found in Appendix B. It characterizes the solutions of (g-BPC) via the signed support of their innovation using the signed saturation set of some dual pre-certificate.

Proposition 1. *Let $\mathbf{x} \in \mathbb{R}^M$ be the ordered sampling locations, and $\mathbf{y}_0 \in \mathbb{R}^M$. An element $f_{\text{opt}} \in \mathrm{BV}^{(2)}(\mathbb{R})$ is a solution of (g-BPC) if and only if f_{opt} satisfies the interpolation conditions $f_{\text{opt}}(x_m) = y_{0,m}$ for all $m \in \{1, \ldots, M\}$ and one can find a dual pre-certificate η (Definition 2) such that*

$$\|w\|_{\mathcal{M}} = \langle w,\ \eta \rangle\,, \tag{16}$$

where $w \overset{\text{def.}}{=} \mathrm{D}^2\{f_{\text{opt}}\}$ is the innovation of f_{opt}. The condition (16) is moreover equivalent to the inclusion

$$\mathrm{supp}_\pm(w) \subset \mathrm{sat}_\pm(\eta). \tag{17}$$

The dual pre-certificate η is then called a dual certificate *(for problem (g-BPC)).*

Remark 1. When $f_{\text{opt}} \in \text{BV}^{(2)}(\mathbb{R})$ is a piecewise-linear spline, *i.e.*, $f_{\text{opt}}(x) = \sum_{k=1}^{K} a_k(x - \tau_k)_+ + b_0 + b_1 x$ for all $x \in \mathbb{R}$ (see (9)), the condition (17) is equivalent to the following interpolation requirements on the dual pre-certificate η

$$\forall k \in \{1, \dots, K\}, \ \eta(\tau_k) = \text{sign}(a_k). \tag{18}$$

From Proposition 1, a dual certificate η is thus a dual pre-certificate that certifies that a given $f_{\text{opt}} \in \text{BV}^{(2)}(\mathbb{R})$ is a solution of (g-BPC), *i.e.*, f_{opt} satisfies $f_{\text{opt}}(x_m) = y_{0,m}$ for all $m \in \{1, \dots, M\}$ and $\text{supp}_{\pm}(\text{D}^2 f_{\text{opt}}) \subset \text{sat}_{\pm}(\eta)$ (or equivalently $\left\|\text{D}^2 f_{\text{opt}}\right\|_{\mathcal{M}} = \left\langle \text{D}^2 f_{\text{opt}}, \ \eta \right\rangle$). Once we know that some η is a dual certificate, it can be used to check whether *any* $f \in \text{BV}^{(2)}(\mathbb{R})$ is a solution of (g-BPC). In other words, contrary to what is seemingly implied in Proposition 1, there is no need to find a new dual pre-certificate for each candidate solution f. This is formulated in the following proposition, the proof of which can be found in Appendix C.

Proposition 2. *Let $\mathbf{x} \in \mathbb{R}^M$ be the ordered sampling locations, $\mathbf{y}_0 \in \mathbb{R}^M$, and let $\eta \in \mathcal{C}_0(\mathbb{R})$ be dual certificate as defined in Proposition 1 for the problem (g-BPC). Then, an element $f_{\text{opt}} \in \text{BV}^{(2)}(\mathbb{R})$ is a solution of (g-BPC) if and only if f_{opt} satisfies the interpolation conditions $f_{\text{opt}}(x_m) = y_{0,m}$ for all $m \in \{1, \dots, M\}$ and*

$$\text{supp}_{\pm}(w) \subset \text{sat}_{\pm}(\eta). \tag{19}$$

or equivalently $\|w\|_{\mathcal{M}} = \langle w, \ \eta \rangle$, where $w \overset{\text{def.}}{=} \text{D}^2 f_{\text{opt}}$ is the innovation of f_{opt}.

To end this section, let us illustrate how the concept of dual certificates can be used to describe the solution set of (g-BPC). Suppose that we know that some η is a dual certificate (we prove in Proposition 5 that this is the case of the dual pre-certificate η_{cano} introduced in Proposition 4), then the condition $\text{supp}_{\pm}(w) \subset \text{sat}_{\pm}(\eta)$ of Proposition 2 enforces strong constraints on any candidate solution of (g-BPC). This is all the more true when $\text{sat}_{\pm}(\eta)$ is a discrete set, which we consider in the next definition and proposition.

Definition 4 (*Nondegeneracy*)**.** Let $\mathbf{x} \in \mathbb{R}^M$ be the ordered sampling locations, $\mathbf{y}_0 \in \mathbb{R}^M$ and let $\eta \in \mathcal{C}_0(\mathbb{R})$ be any dual certificate as defined in Proposition 1. We say that η is *nondegenerate* if its signed saturation set $\text{sat}_{\pm}(\eta)$ defined in Definition 3 is a discrete set. Otherwise, we say that it is *degenerate*.

Proposition 3 (*General Uniqueness Result for (g-BPC)*)**.** *Let $\mathbf{x} \in \mathbb{R}^M$ be the ordered sampling locations and $\mathbf{y}_0 \in \mathbb{R}^M$. If there exists a nondegenerate dual certificate in the sense of Definition 2, then the optimization problem (g-BPC) has a unique solution, which is a piecewise-linear spline in the sense of Definition 1 with $K \leq M - 2$ knots τ_k that form a subset of the sampling points $\{x_2, \dots, x_{M-1}\}$.*

The proof of Proposition 3 is given in Appendix D.

3. The solutions of the (g-BPC)

In this section, we consider the optimization problem (g-BPC) where the x_m for $m \in \{1, \dots, M\}$ are distinct and ordered sampling locations and $\mathbf{y}_0 \in \mathbb{R}^M$ is a fixed measurement vector. This setting is especially relevant when the measurements $y_{0,m}$ are exactly the values of the input signal at locations x_m (noiseless case). The solution set is

$$\mathcal{V}_0 \overset{\text{def.}}{=} \underset{\substack{f \in \text{BV}^{(2)}(\mathbb{R}) \\ f(x_m) = y_{0,m}, \ m \in \{1, \dots, M\}}}{\arg\min} \|\text{D}^2 f\|_{\mathcal{M}}, \tag{g-BPC}$$

and is known to admit at least one piecewise-linear solution due to Theorem 1.

3.1. Canonical solution and canonical dual certificate

Thereafter, we identify the complete set of solutions (g-BPC). This allows us to fully determine in which cases this optimization problem admits a unique solution. Our analysis is based on the construction of a pair $(f_{\text{cano}}, \eta_{\text{cano}}) \in \text{BV}^{(2)}(\mathbb{R}) \times \mathcal{C}_0(\mathbb{R})$ that satisfies Proposition 1, which we call the canonical solution and canonical dual certificate respectively. The former is simply the function that connects the points $\text{P}_{0,m} = \begin{bmatrix} x_m & y_{0,m} \end{bmatrix}^T$.

Definition 5 (*Canonical Interpolant*)**.** Let $\mathbf{x} \in \mathbb{R}^M$ be the ordered sampling locations and $\mathbf{y}_0 \in \mathbb{R}^M$ with $M \geq 2$. We define f_{cano} as the unique piecewise-linear spline that interpolates the data points with the minimum number of knots, *i.e.*, such that

- $f_{\text{cano}}(x_m) = y_{0,m}$ for any $m \in \{1, \dots, M\}$ and
- f_{cano} has at most $M - 2$ knots which form a subset of $\{x_m : 2 \leq m \leq M - 1\}$.

We refer to f_{cano} as the *canonical interpolant*.

The existence and uniqueness of f_{cano} in Definition 5 simply follows from the number of degrees of freedom of a piecewise-linear spline whose knots are known. The canonical interpolant is of the form

$$f_{\text{cano}}(x) = a_1 x + a_M + \sum_{m=2}^{M-1} a_m (x - x_m)_+ \tag{20}$$

with $\mathbf{a} = (a_1, \dots, a_M) \in \mathbb{R}^M$. By definition, f_{cano} is linear on the interval (x_m, x_{m+1}) for $m \in \{2, \dots, M-1\}$. The interpolatory conditions $f_{\text{cano}}(x_m) = y_m$ and $f_{\text{cano}}(x_{m+1}) = y_{m+1}$ then imply that its slope is $s_m = \frac{y_{0,m+1} - y_{0,m}}{x_{m+1} - x_m}$. Yet from (20) we get that $s_m = a_1 + \cdots + a_m$. This implies that $a_1 = s_1$ and that $a_m = s_m - s_{m-1}$ for $m \in \{2, \dots, M-1\}$. Finally, the equation $f_{\text{cano}}(x_1) = y_{0,1}$ yields $a_M = y_{0,1} - a_1 x_1$. Consequently, the vector $\mathbf{a} \in \mathbb{R}^M$ in (20) is given by

$$\begin{cases} a_1 = \frac{y_{0,2} - y_{0,1}}{x_2 - x_1}, \\ a_m = \frac{y_{0,m+1} - y_{0,m}}{x_{m+1} - x_m} - \frac{y_{0,m} - y_{0,m-1}}{x_m - x_{m-1}}, \quad \forall m \in \{2, \dots, M-1\}, \\ a_M = y_{0,1} - \frac{y_{0,2} - y_{0,1}}{x_2 - x_1} x_1. \end{cases} \tag{21}$$

In order to prove that f_{cano} is always a solution of (g-BPC), we construct a particular dual pre-certificate η_{cano}.

Proposition 4 (*Canonical Pre-Certificate*). *Let $\mathbf{x} \in \mathbb{R}^M$ be the ordered sampling locations and $\mathbf{y}_0 \in \mathbb{R}^M$. Let $\mathbf{a} \in \mathbb{R}^M$ be the vector defined by (21). There exists a unique piecewise-linear spline η_{cano} given by*

$$\eta_{\text{cano}} \overset{\text{def.}}{=} \sum_{m=1}^{M} c_m (x_m - \cdot)_+ \quad \text{with} \quad \mathbf{c} = (c_1, \dots, c_M) \in \mathbb{R}^M, \tag{22}$$

$$\langle \mathbf{c}, \mathbf{1} \rangle = \langle \mathbf{c}, \mathbf{x} \rangle = 0, \tag{23}$$

$$\forall m \in \{2, \dots, M-1\}, \ \eta_{\text{cano}}(x_m) = \operatorname{sign}(a_m). \tag{24}$$

with the convention $\operatorname{sign}(0) = 0$. *Moreover, since $\eta_{\text{cano}}(x) = 0$ for $x \le x_1$ and $x \ge x_M$, we have $\eta_{\text{cano}} \in \mathcal{C}_0(\mathbb{R})$ and $\|\eta_{\text{cano}}\|_\infty = 1$. Hence, η_{cano} is a dual pre-certificate in the sense of Definition 2.*

Proof. The existence and uniqueness of such a spline follows the same argument as for f_{cano}, applied to the data points $(x_1 - 1, 0)$, $(x_1, 0)$, $(x_m, \operatorname{sign}(a_m))$ for $m \in \{2, \dots, M-1\}$, $(x_M, 0)$ and $(x_M + 1, 0)$. Note that the points $(x_1 - 1, 0)$ and $(x_M + 1, 0)$ at the boundaries add two additional interpolation constraints to (24). Moreover, they imply that η_{cano} does not have a linear term and is thus of the form (22).

Next, we notice that for $x \le x_1$, we have $\eta_{\text{cano}}(x) = -\langle \mathbf{c}, \mathbf{x} \rangle x + \langle \mathbf{c}, \mathbf{1} \rangle = 0$, due to $\langle \mathbf{c}, \mathbf{x} \rangle = \langle \mathbf{c}, \mathbf{1} \rangle = 0$. For $x \ge x_M$, $(x_m - x)_+ = 0$ for every $m \in \{1, \dots, M\}$, hence $\eta_{\text{cano}}(x) = 0$. Then, as a piecewise-linear spline with compact support, η_{cano} is of course in $\mathcal{C}_0(\mathbb{R})$. Being compactly supported, it is also clear that η_{cano} attains its maximum and minimum values at its knots. In particular, $\|\eta_{\text{cano}}\|_\infty = \max_{m \in \{1, \dots, M\}} |\eta_{\text{cano}}(x_m)| = 1$.

We now prove that the pair $(f_{\text{cano}}, \eta_{\text{cano}}) \in \mathrm{BV}^{(2)}(\mathbb{R}) \times \mathcal{C}_0(\mathbb{R})$ satisfies Proposition 1. Although the fact that f_{cano} is a solution to (g-BPC) is known [16,17] and is significant in its own right, the key element of this result is the construction of the dual certificate η_{cano}. The latter will be essential to fully describe the solution set $\mathcal{V}_0$.

Proposition 5. *Let $\mathbf{x} \in \mathbb{R}^M$ be the ordered sampling locations and $\mathbf{y}_0 \in \mathbb{R}^M$. The canonical interpolant f_{cano} defined in Definition 5 is a solution of (g-BPC) and η_{cano}, defined in Proposition 4, is a dual certificate as defined in Proposition 1.*

Proof. By construction, the interpolation conditions $f_{\text{cano}}(x_m) = y_{0,m}$ for all $m \in \{1, \dots, M\}$ are satisfied. Moreover thanks to Proposition 4, η_{cano} is a dual pre-certificate. By Proposition 1, it remains to prove that

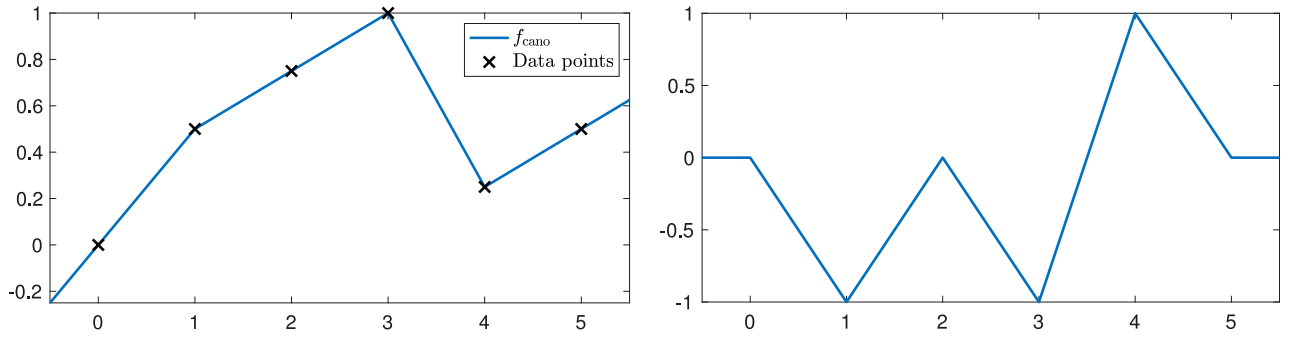
$$\operatorname{supp}_\pm(\mathrm{D}^2 f_{\text{cano}}) \subset \operatorname{sat}_\pm(\eta_{\text{cano}}), \tag{25}$$

from which we deduce both that f_{cano} is a solution of (g-BPC) and that η_{cano} is a dual certificate. Since, again by construction, $\eta_{\text{cano}}(x_m) = \operatorname{sign}(a_m)$ for all $m \in \{1, \dots, M\}$ and $\mathrm{D}^2 f_{\text{cano}} = \sum_{m=2}^{M-1} a_m \delta(\cdot - x_m)$, this proves (25).

Due to Proposition 5, we call f_{cano} the *canonical solution* and η_{cano} the *canonical dual certificate* of the optimization problem (g-BPC). We show an example of such functions for given data points $(x_m, y_{0,m})_{m \in \{1, \dots, 6\}}$ in Fig. 2. Notice that the points $\mathrm{P}_{0,2}$, $\mathrm{P}_{0,3}$, and $\mathrm{P}_{0,4}$ are aligned, which implies that $a_3 = 0$ (defined in (21)).

3.2. Characterization of the solution set

Although identifying a solution f_{cano} to (g-BPC) is an important first step, this solution is not unique in general. We characterize the case of uniqueness in Proposition 6, and then provide a complete description of the solution set when the solution is not unique in Theorem 1. We shall see that the canonical dual certificate η_{cano} plays an essential role regarding these issues.



(a) Canonical solution f_{cano} (b) Canonical dual certificate η_{cano}

Fig. 2. Example of a canonical solution and canonical dual certificate for $M = 6$ with $x_m = m - 1$. We have $a_2 < 0$, $a_3 = 0$, $a_4 < 0$, and $a_5 > 0$, where the a_m are defined in (21).

Proposition 6 (*Uniqueness Result for* (g-BPC))**.** *Let* $\mathbf{x} \in \mathbb{R}^M$ *be the ordered sampling locations and* $\mathbf{y}_0 \in \mathbb{R}^M$*. Then, the following conditions are equivalent.*

1. (g-BPC) *has a unique solution.*
2. *The canonical dual certificate* η_{cano} *(defined in Proposition 4) is nondegenerate (see Definition 4).*
3. *For all* $m \in \{2, \dots, M-2\}$, $a_m a_{m+1} \leq 0$*, where* $\mathbf{a} \in \mathbb{R}^M$ *is given by* (21)*.*

Proof. The equivalence 2. $\Leftrightarrow$ 3. comes from the fact that η_{cano} is nondegenerate if and only if it never saturates at 1 or -1 between two consecutive knots. This is equivalent to item 3 because for all $m \in \{2, \dots, M-1\}$, $\eta_{\text{cano}}(x_m) = \text{sign}(a_m)$.

The implication 2. $\Rightarrow$ 1. is given by Proposition 3. We now prove the contraposition of the reverse implication 1. $\Rightarrow$ 2. We thus assume that η_{cano} is degenerate, and wish to prove that (g-BPC) has multiple solutions. Using item 3., there exists an index $m \in \{2, \dots, M-2\}$ such that $a_m a_{m+1} > 0$. We now invoke the following lemma (illustrated in Fig. 3) that plays an important role throughout the paper.

Lemma 1. *Let* $\mathbf{x} \in \mathbb{R}^M$ *be the ordered sampling locations, and* $\mathbf{y}_0 \in \mathbb{R}^M$ *with* $M \geq 4$*. Let* $m \in \{2, \dots, M-2\}$ *be an index such that* $a_m a_{m+1} > 0$*, where* $\mathbf{a} \in \mathbb{R}^M$ *is defined as in* (21)*. Then, the lines* $(\text{P}_{0,m-1}, \text{P}_{0,m})$ *and* $(\text{P}_{0,m+1}, \text{P}_{0,m+2})$ *are intersecting at a point* $\widetilde{\text{P}} = \begin{bmatrix} \tilde{\tau} & \tilde{y} \end{bmatrix}^T$ *such that* $x_m < \tilde{\tau} < x_{m+1}$*. Moreover, the piecewise-linear spline* f_{opt} *defined by*

$$f_{\text{opt}}(x) \overset{\text{def.}}{=} \begin{cases} \frac{y_{0,m} - y_{0,m-1}}{x_m - x_{m-1}}(x - x_{m-1}) + y_{0,m-1}, & \text{for } x_m < x \leq \tilde{\tau} \\ \frac{y_{0,m+2} - y_{0,m+1}}{x_{m+2} - x_{m+1}}(x - x_{m+1}) + y_{0,m+1}, & \text{for } \tilde{\tau} < x < x_{m+1} \\ f_{\text{cano}}(x) & \text{for } x \notin (x_m, x_{m+1}), \end{cases} \tag{26}$$

which has no knots at x_m *or* x_{m+1}*, is a solution of* (g-BPC)*.*

Proof. Let $I_0 = \{2, \dots, M-1\} \setminus \{m, m+1\}$. We then define

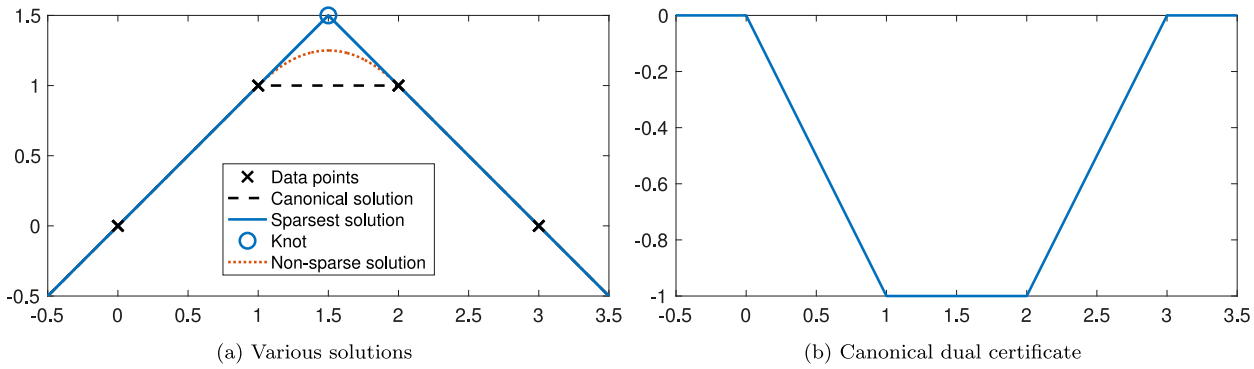
$$f_{\text{opt}}(x) \overset{\text{def.}}{=} a_1 x + a_M + \sum_{m' \in I_0} a_{m'}(x - x_{m'})_+ + \tilde{a}(x - \tilde{\tau})_+, \tag{27}$$

where $\tilde{a} = a_m + a_{m+1}$ and $\tilde{\tau} = \frac{a_m x_m + a_{m+1} x_{m+1}}{\tilde{a}}$. By definition, $\tilde{\tau}$ is a barycenter of x_m and x_{m+1} with weights $\frac{a_m}{\tilde{a}}$ and $\frac{a_{m+1}}{\tilde{a}}$. Yet a_m and a_{m+1} have the same (nonzero) signs, which implies that these weights are in the interval $(0, 1)$ and thus that $\tilde{\tau} \in (x_m, x_{m+1})$. Yet f_{opt} has no knots at x_m and x_{m+1}, so it must follow the line $(\text{P}_{0,m-1}, \text{P}_{0,m})$ in the interval $[x_m, \tilde{\tau}]$, and the line $(\text{P}_{0,m+1}, \text{P}_{0,m+2})$ in the interval $[\tilde{\tau}, x_{m+1}]$, which conforms with the first two first lines in (26). Due to the continuity of f_{opt}, these lines are therefore intersecting at the point $\widetilde{\text{P}} = \begin{bmatrix} \tilde{\tau} & \tilde{y} \end{bmatrix}^T = \begin{bmatrix} \tilde{\tau} & f_{\text{opt}}(\tilde{\tau}) \end{bmatrix}^T$.

Next, for $x \leq x_m$, we have $a_m(x - x_m)_+ + a_{m+1}(x - x_{m+1})_+ = \tilde{a}(x - \tilde{\tau})_+ = 0$. Similarly, for $x \geq x_{m+1}$, we have $a_m(x - x_m)_+ + a_{m+1}(x - x_{m+1})_+ = \tilde{a}(x - \tilde{\tau})_+ = \tilde{a}(x - \tilde{\tau})$ since $x \geq \tilde{\tau}$. Therefore, for any $x \notin (x_m, x_{m+1})$, we have $f_{\text{cano}}(x) = f_{\text{opt}}(x)$, which conforms with the third line in (26). This also implies that $f_{\text{opt}}(x_m) = f_{\text{cano}}(x_m) = y_{0,m}$ for all $m \in \{1, \dots, M\}$. Moreover, we have $\|\text{D}^2 f_{\text{cano}}\|_{\mathcal{M}} = \sum_{m=2}^{M-1} a_m = \sum_{m \in I_0} a_m + \tilde{a} = \|\text{D}^2 f_{\text{opt}}\|_{\mathcal{M}}$. Therefore, f_{opt} has the same measurements and regularization cost as f_{cano}, which implies that it is also a solution of (g-BPC).

Since f_{opt} defined in Lemma 1 is a solution to the (g-BPC) such that $f_{\text{opt}} \neq f_{\text{cano}}$, the (g-BPC) has multiple solutions, which concludes the proof.

To the best of our knowledge, Proposition 6 is a new result. A similar uniqueness result is presented in [60, Theorem 4.2], but with more restrictive conditions than item 3. It follows from Proposition 6 that when $M = 3$, the solution of



(a) Various solutions (b) Canonical dual certificate

Fig. 3. Example with $M = 4$ of a non-unique solution (η_{cano} saturates at -1). An example of a non-sparse solution with a quadratic regime in $[1, 2]$ is given.

the (g-BPC) is always unique because the certificate is always nondegenerate, and is given by f_{cano}. We go much further in Theorem 2 by providing the full characterization of the solution set when $M \geq 4$.

Theorem 2 (*Characterization of the Solution Set of the* (g-BPC)). *Let* $\mathbf{x} \in \mathbb{R}^M$ *be the ordered sampling locations and* $\mathbf{y}_0 \in \mathbb{R}^M$ *with* $M \geq 4$, *and let* f_{cano} *and* η_{cano} *be the functions defined in Definition 5 and Proposition 4 respectively. A function* $f_{\text{opt}} \in \text{BV}^{(2)}(\mathbb{R})$ *is a solution of the* (g-BPC) *if and only if* $f_{\text{opt}}(x_m) = y_{0,m}$ *for* $m \in \{1, \dots, M\}$, *and the following conditions are satisfied for* $m \in \{2, \dots, M-2\}$

1. $f_{\text{opt}} = f_{\text{cano}}$ *in* $[x_m, x_{m+1}]$ *if* $|\eta_{\text{cano}}| < 1$ *in* (x_m, x_{m+1});
2. f_{opt} *is convex in* $[x_{m-1}, x_{m+2}]$ *if* $\eta_{\text{cano}} = 1$ *in* $[x_m, x_{m+1}]$;
3. f_{opt} *is concave in* $[x_{m-1}, x_{m+2}]$ *if* $\eta_{\text{cano}} = -1$ *in* $[x_m, x_{m+1}]$;
4. $f_{\text{opt}} = f_{\text{cano}}$ *in* $(-\infty, x_2)$ *and* $(x_{M-1}, +\infty)$.

Proof. Let f_{opt} be a solution of the (g-BPC). According to Proposition 5, η_{cano} is a dual certificate. According to Proposition 2, we therefore have that $\operatorname{supp}_{\pm}(\mathrm{D}^2 f_{\text{opt}}) \subset \operatorname{sat}_{\pm}(\eta_{\text{cano}})$, meaning that $\mathrm{D}^2 f_{\text{opt}} = 0$ on the complement $\operatorname{sat}_{\pm}(\eta_{\text{cano}})^c$ of $\operatorname{sat}_{\pm}(\eta_{\text{cano}})$. In particular, we have that $(-\infty, x_2] \subset \operatorname{sat}_{\pm}(\eta_{\text{cano}})^c$, hence f_{opt} is linear on this interval. The interpolation constraints $f_{\text{opt}}(x_1) = f_{\text{cano}}(x_1)$ and $f_{\text{opt}}(x_2) = f_{\text{cano}}(x_2)$ then imply that $f_{\text{opt}} = f_{\text{cano}}$ on $(-\infty, x_2]$. The same argument holds for the interval $[x_{M-1}, +\infty)$ and any interval (x_m, x_{m+1}) on which η_{cano} does not saturate.

Assume now that $[x_m, x_{m+1}] \subset \operatorname{sat}_{+}(\eta_{\text{cano}})$; that is, $\eta_{\text{cano}} = 1$ on $[x_m, x_{m+1}]$. We use the Jordan decomposition of $\mathrm{D}^2 f_{\text{opt}} = w = w_+ - w_-$ where w_+ and w_- are positive measures. By (17), we know that $w_- = 0$ on $[x_m, x_{m+1}]$ because its support is included in $\operatorname{sat}_{-}(\eta_{\text{cano}})$. Hence, on this interval, $\mathrm{D}^2 f_{\text{opt}} = w = w_+$ is a positive measure, implying that $\mathrm{D} f_{\text{opt}}$ is increasing and therefore that f_{opt} is convex on $[x_m, x_{m+1}]$. Now, if $(x_{m-1}, x_m) \subset \operatorname{sat}_{+}(\eta)^c \cap \operatorname{sat}_{-}(\eta)^c$ then, as above, $\mathrm{D}^2 f_{\text{opt}|(x_{m-1}, x_m)} = 0$. Otherwise, by continuity of η_{cano}, we have $(x_{m-1}, x_m) \subset \operatorname{sat}_{+}(\eta_{\text{cano}})$ hence $\mathrm{D}^2 f_{\text{opt}|(x_{m-1}, x_m)} \geq 0$. As a result f_{opt} is convex on $(x_{m-1}, x_{m+1}]$. The same argument proves that f_{opt} is convex on $[x_m, x_{m+2})$, and therefore on the whole interval (x_{m-1}, x_{m+2}).

Suppose conversely that f_{opt} satisfies all the conditions of Theorem 2. Let us prove that it is a solution of the (g-BPC). By Proposition 2, we just need to check that f_{opt} satisfies $\operatorname{supp}_{\pm}(\mathrm{D}^2 f_{\text{opt}}) \subset \operatorname{sat}_{\pm}(\eta_{\text{cano}})$ since by construction, $f_{\text{opt}}(x_m) = y_{0,m}$. By definition of η_{cano}, we have $\mathrm{D}^2 f_{\text{opt}} = 0$ on $\operatorname{sat}_{+}(\eta_{\text{cano}})^c \cap \operatorname{sat}_{-}(\eta_{\text{cano}})^c$ (because $\mathrm{D}^2 f_{\text{opt}}$ is equal to f_{cano} which is linear on that set). Moreover, $\mathrm{D}^2 f_{\text{opt}} \geq 0$ on $\operatorname{sat}_{+}(\eta_{\text{cano}})$ (because by assumption, f_{opt} is convex on intervals where $\eta_{\text{cano}} = 1$) and $\mathrm{D}^2 f_{\text{opt}} \leq 0$ on $\operatorname{sat}_{-}(\eta_{\text{cano}})$ (because f_{opt} is concave on intervals where $\eta_{\text{cano}} = -1$). This means that $\operatorname{supp} w_+ \subset \operatorname{sat}_{+}(\eta_{\text{cano}})$ and $\operatorname{supp} w_- \subset \operatorname{sat}_{+}(\eta_{\text{cano}})$ where $\mathrm{D}^2 f_{\text{opt}} = w_+ - w_-$ is again the Jordan decomposition of $\mathrm{D}^2 f_{\text{opt}}$. Finally, as expected, we have that

$$\operatorname{supp}_{\pm}(\mathrm{D}^2 f_{\text{opt}}) = \operatorname{supp} w_+ \times \{1\} \cup \operatorname{supp} w_- \times \{-1\} \subset \operatorname{sat}_{+}(\eta_{\text{cano}}) \times \{1\} \cup \operatorname{sat}_{-}(\eta_{\text{cano}}) \times \{-1\} = \operatorname{sat}_{\pm}(\eta_{\text{cano}}), \tag{28}$$

hence f_{opt} is a solution of the (g-BPC).

To illustrate Theorem 2, a simple example with $M = 4$ data points for which the solution is not unique is given in Theorem 2. Indeed, the canonical dual certificate saturates at -1 in the interval $[1, 2]$. Therefore, by Theorem 2, any function that coincides with f_{cano} in $\mathbb{R} \setminus [1, 2]$ and that is concave in the interval $[0, 3]$ is a solution. This includes the sparsest solution (with a single knot), as well as non-sparse solutions, *e.g.*, with a quadratic regime in $[1, 2]$ as in Fig. 3.

Corollary 1. *If the* (g-BPC) *has more than one solution, then it has an uncountable number of solutions.*

Proof. If the solution is not unique, then the dual certificate η_{cano} is degenerate, and therefore saturates over some interval (x_m, x_{m+1}). Then, Theorem 2 characterizes the whole set of solutions, which is clearly uncountably infinite.

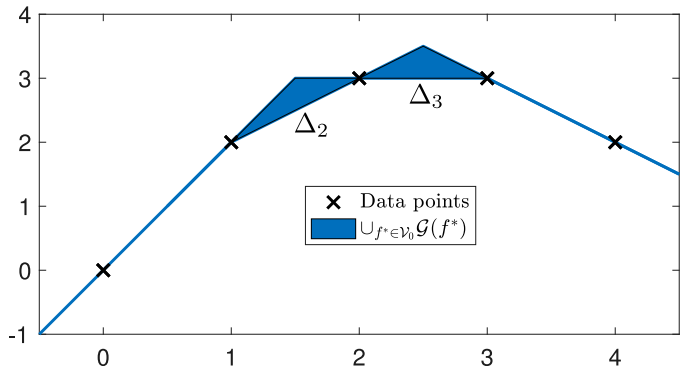


Fig. 4. Example with $M = 5$ of the geometric domain $\cup_{f_{\text{opt}} \in \mathcal{V}_0} \mathcal{G}(f_{\text{opt}})$ containing all the solutions to the (g-BPC). We have $\mathcal{X} = \{2, 3\}$ and thus two triangles Δ_m; all solutions follow f_{cano} everywhere else.

Corollary 1 is the continuous counterpart of the well-known fact that the discrete LASSO either admits a unique solution or an uncountable number of solutions [89, Lemma 1]. Even with infinitely many solutions, we are able to delimit the geometric domain that contains the graphs of all solutions by exploiting the local convex/concavity. We recall that $\mathrm{P}_{0,m} = [x_m \; y_{0,m}]^T$ for $m \in \{1, \dots, M\}$, and that for $\mathrm{A}, \mathrm{B} \in \mathbb{R}^2$, we denote by (A, B) the line joining A and B. Then, for $M \geq 4$, we consider the set of indices

$$\mathcal{X} \overset{\text{def.}}{=} \mathcal{X}(\mathbf{x}, \mathbf{y}_0) \overset{\text{def.}}{=} \{m \in \{2, \dots, M-2\};\; a_m a_{m+1} > 0\}\,, \tag{29}$$

where we recall that $a_m = \frac{y_{0,m+1} - y_{0,m}}{x_{m+1} - x_m} - \frac{y_{0,m} - y_{0,m-1}}{x_m - x_{m-1}}$ (see (21)). The slope condition $a_m a_{m+1} > 0$ in (29) is equivalent to the fact that the lines $(\mathrm{P}_{0,m-1}, \mathrm{P}_{0,m})$ and $(\mathrm{P}_{0,m+1}, \mathrm{P}_{0,m+2})$ are not parallel (otherwise we would have that $a_m = -a_{m+1}$, hence $a_m a_{m+1} \leq 0$) and that their intersection point, that we denote by $\widetilde{\mathrm{P}}_m = [\tilde{\tau}_m \; \tilde{y}_m]^T$, is such that $x_m \leq \tilde{\tau}_m \leq x_{m+1}$ according to Lemma 1. We can thus introduce the triangles Δ_m, whose vertices are the points $\mathrm{P}_{0,m}$, $\widetilde{\mathrm{P}}_m$, and $\mathrm{P}_{0,m+1}$. Theorem 3 makes the link between the graph of any solution $f_{\text{opt}} \in \mathrm{BV}^{(2)}(\mathbb{R})$ of the (g-BPC), the graph of f_{cano} and the triangles Δ_m.

Theorem 3 (*Geometric Domain of the Graph of Solutions of the* (g-BPC)). *Let* $\mathbf{x} \in \mathbb{R}^M$ *be the ordered sampling locations and* $\mathbf{y}_0 \in \mathbb{R}^M$ *with* $M \geq 4$. *Then, we have*

$$\cup_{f_{\text{opt}} \in \mathcal{V}_0} \mathcal{G}(f_{\text{opt}}) = \mathcal{G}(f_{\text{cano}}) \cup (\cup_{m \in \mathcal{X}} \Delta_m)\,, \tag{30}$$

where f_{cano} *is defined in Definition 5,* $\mathcal{X}$ *is defined in* (29)*, and the* Δ_m *triangles are defined in the above paragraph.*

The relation (30) reveals the smallest possible geometric domain containing all the graphs of the solutions of the (g-BPC). To obtain a solution of the (g-BPC), one just needs to follow the graph of f_{cano} outside the triangles Δ_m and take a convex or concave function inside them. An example of this domain is given in Fig. 4 with $M = 5$ and $\#\mathcal{X} = 2$ triangles (this same example is treated further later in Fig. 6). The proof of Theorem 3 is given in Appendix E. Next, Section 4 is dedicated to the study of the sparsest piecewise-linear solutions of the (g-BPC).

4. The sparsest solution(s) of the (g-BPC)

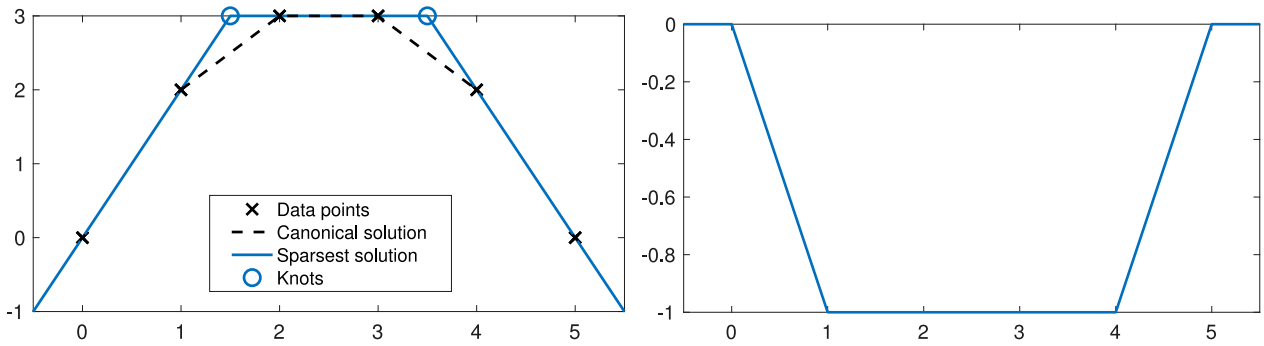
4.1. Characterization of the sparsest solution(s)

We have already identified the situations where the (g-BPC) admits a unique solution, in which case it is the canonical solution introduced in Definition 5. When the solution is not unique, Theorem 1 ensures that the extreme-point solutions are piecewise-linear functions with at most $K - 2$ knots, and Theorem 2 gives a complete description of the solution set. In this section, we go further by providing a complete answer to the following questions:

- what is the minimal number of knots of a solution of the (g-BPC)?
- what are the sparsest solutions, *i.e.*, the ones reaching this minimum number of knots?

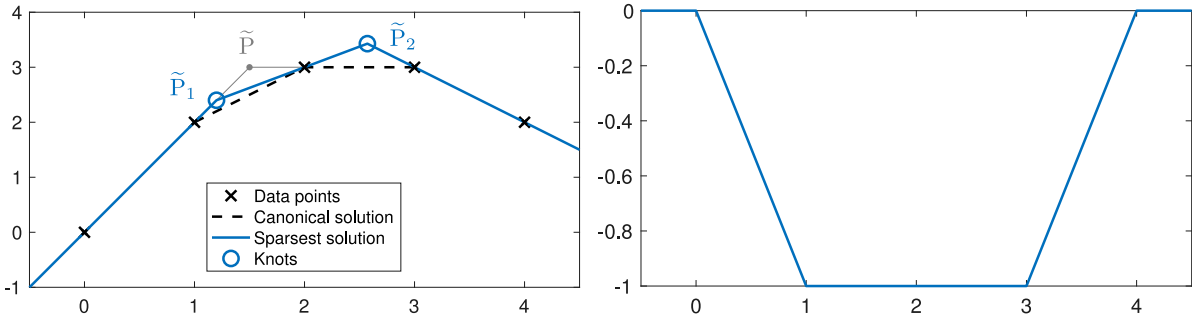
These questions are addressed in Theorem 4. Let η_{cano} be defined as in Proposition 4 for fixed values of $\mathbf{x}, \mathbf{y}_0 \in \mathbb{R}^M$, and let

$$\begin{aligned} I_{\text{sat}} &\overset{\text{def.}}{=} \{m \in \{2, \dots, M-1\} : \eta_{\text{cano}}(x_m) = \pm 1 \text{ and } \eta_{\text{cano}}(x_m) \neq \eta_{\text{cano}}(x_{m-1})\} \\ &= \{s_1, \dots, s_{N_s}\} \quad \text{with} \quad s_1 < \cdots < s_{N_s}. \end{aligned} \tag{31}$$



(a) Sparsest solution (b) Canonical certificate

Fig. 5. Example with $M = 6$ and $\alpha = 3$ consecutive saturation intervals of η_{cano} at -1. The unique sparsest solution has $P = 2$ knots.



(a) Example of a sparsest solution (b) Canonical certificate

Fig. 6. Example with $M = 5$ and $\alpha = 2$ consecutive saturation intervals of η_{cano} at -1. The sparsest solutions have $P = 2$ knots.

In other words, $N_s = \#I_{\text{sat}}$ corresponds to the number of times η_{cano} reaches ± 1. Next, let $\alpha_n \in \mathbb{N}$ for $n \in \{1, \dots, N_s\}$ be the number of consecutive intervals starting from x_{s_n} in which η_{cano} saturates at ± 1, *i.e.*,

$$\alpha_n \overset{\text{def.}}{=} \min\{k \in \mathbb{N} : \eta_{\text{cano}}(x_{s_n+k+1}) \neq \eta_{\text{cano}}(x_{s_n})\}. \tag{32}$$

In what follows, $\lceil x \rceil$ is the smallest integer larger than or equal to $x \in \mathbb{R}$.

Theorem 4 (*Sparsest Solutions of the* (g-BPC)). *Let* $\mathbf{x} \in \mathbb{R}^M$ *be the ordered sampling locations,* $\mathbf{y}_0 \in \mathbb{R}^M$ *with* $M \geq 4$. *Concerning the minimum sparsity of a solution of the* (g-BPC), *the following hold.*

1. *The lowest possible sparsity* (i.e., *number of knots) of a piecewise-linear solution of the* (g-BPC) *is*

$$K_{min}(\mathbf{x}, \mathbf{y}_0) = \sum_{n=1}^{N_s} \left\lceil \frac{\alpha_n + 1}{2} \right\rceil, \tag{33}$$

   where the α_n *are defined in* (32), *and* $N_s = \#I_{sat}$ *where* I_{sat} *is defined in* (31).
2. *There is a unique sparsest solution of the* (g-BPC) *if and only if none of the* α_n *are nonzero even numbers.*
3. *If one or more* $\alpha_n > 0$ *are even, then there are uncountably many sparsest solutions to the* (g-BPC). *The number of degrees of freedom* $n_{\text{free}}(\mathbf{x}, \mathbf{y}_0)$ *of the set of sparsest solutions is equal to the number of even* α_n *coefficients, that is,*

$$n_{\text{free}}(\mathbf{x}, \mathbf{y}_0) = \sum_{n=1}^{N_s} \mathbb{1}_{2\mathbb{N}_{\geq 1}}(\alpha_n). \tag{34}$$

   More precisely, for each saturation region of η_{cano}, *fixing a single knot within a certain admissible segment uniquely determines the other knots within the saturation region.*

The proof of Theorem 4 is given in Appendix F. Illustrations of its items 2. and 3. with a single saturation region (*i.e.*, $N_s = 1$) are given in Figs. 5 and 6 respectively. In Fig. 5, the unique sparsest solution is shown. In Fig. 6, any point $\widetilde{\mathrm{P}}_1$ in the segment that connects the points $\mathrm{P}_{0,2}$ and $\widetilde{\mathrm{P}}$ yields one of the sparsest solutions, with a uniquely determined second knot $\widetilde{\mathrm{P}}_2$. In the latter example, there is thus a single degree of freedom $n_{\text{free}}(\mathbf{x}, \mathbf{y}_0)$ in the set of sparsest solutions to the (g-BPC).

Input: $\mathbf{x}$, $\mathbf{y}_0$
compute $a_1, \dots a_M$ defined in Eq. (21); $[\eta_{\text{cano}}(x_1), \dots, \eta_{\text{cano}}(x_M)] = [0, \text{sign}(a_2), \dots, \text{sign}(a_{M-1}), 0]$;
compute N_s, $s_1, \dots, s_{N_s}$ and $\alpha_1, \dots, \alpha_{N_s}$ defined in Eq. (31) and Eq. (32);
$\hat{\boldsymbol{\tau}} = [\,]$; $\hat{\mathbf{a}} = [\,]$;
for $n \leftarrow 1$ **to** N_s **do**
  $P \leftarrow \lceil \frac{\alpha_n+1}{2} \rceil$;
  compute $\tilde{\tau}_1, \dots, \tilde{\tau}_P$ and $\tilde{a}_1, \dots, \tilde{a}_P$ using Eq. (F.3) or Eq. (F.4);
  $\hat{\boldsymbol{\tau}} \leftarrow [\hat{\boldsymbol{\tau}}, \tilde{\tau}_1, \dots, \tilde{\tau}_P]$;
  $\hat{\mathbf{a}} \leftarrow [\hat{\mathbf{a}}, \tilde{a}_1, \dots, \tilde{a}_P]$;
end
return $f_{\text{opt}} \leftarrow \sum_{k=1}^{K} \hat{a}_k(\cdot - \hat{\tau}_k)_+$

Algorithm 1: Pseudocode of our algorithm to find a sparsest solution of the Eq. (g-BPC).

4.2. Algorithm for reaching a sparsest solution

The results of Theorem 4 suggest a simple yet elegant algorithm for constructing a sparsest solution of the (g-BPC) for given sampling locations $\mathbf{x} = (x_1, \dots, x_M)$ and data $\mathbf{y}_0 = (y_{0,1}, \dots, y_{0,M})$. The pseudocode is given in Algorithm 1, which applies the sparsifying procedure described in Lemma 6 in every saturation interval. Since the latter is rather lengthy and technical, it is given in Appendix F for ease of reading. The proof of Theorem 4 guarantees that the output f^* of Algorithm 1 is indeed a sparsest solution to the (g-BPC), with sparsity $K_{\min}(\mathbf{x}, \mathbf{y}_0)$ as defined in Theorem 4. The following observations can be made concerning Algorithm 1.

- In the cases where the sparsest solution is not unique, the choice of solution specified by (F.4) (which is *not* the one shown in Fig. 6) is guided by simplicity. However, it is an arbitrary choice that can be adapted depending on the application.
- Notice that the x_m such that $\eta_{\text{cano}}(x_m) = 0$ need not be included in the vector of knots $\mathbf{x}'$ built in the algorithm, since we have $a_m = 0$. Therefore, there is in fact no knot at x_m in the canonical solution, which implies that the sparsity of f_{cano} is strictly less than $M - 2$. This corresponds to alignment cases of the data points, *i.e.*, the points $\mathrm{P}_{0,m-1}$, $\mathrm{P}_{0,m}$, and $\mathrm{P}_{0,m+1}$ are aligned, as illustrated in Fig. 2.
- Algorithm 1 can be translated into an online algorithm, *i.e.*, an updated solution can be computed efficiently if a new input data point is added. More precisely, when a new data point $\mathrm{P}_{0,M+1}$ is added, the reconstructed signal is at worst only modified in the saturation interval $I = [x_{s_n-1}, x_{s_n+\alpha_n}]$ if $x_{M+1} \in I$. Since in practice, we usually have $\alpha_n \ll M$, the computational complexity of updating the solution is typically much smaller than rerunning the complete offline algorithm.

4.3. Computational complexity

Algorithm 1 is very fast and memory-efficient; it requires at most two passes through the data points, and thus has linear time and space complexity $\mathcal{O}(M)$ with respect to the number of data points. More precisely, computing the canonical interpolant (*i.e.*, , the a_m coefficients using (21)) requires about $3M$ subtractions and M divisions, and storing two arrays of size M. Next, in the worst-case scenario where $\text{sign}(a_2) = \cdots = \text{sign}(a_{M-1})$, computing the sparsest interpolant (*i.e.*, the $\tilde{a}_k$ and $\tilde{x}_k$ coefficients using (F.3) or (F.4)) requires approximately M multiplications, M additions, $\frac{M}{2}$ divisions and storing two arrays of size $\frac{M}{2}$. Hence, the complete worst-case time complexity for Algorithm 1 requires $4M$ additions, M multiplications and $\frac{3M}{2}$ divisions, and its space complexity is $3M$.

5. The solutions of the (g-BLASSO)

We now focus on the (g-BLASSO) problem, in which the interpolation of the data is no longer required to be exact as in Section 3, but is formulated as a penalized problem with a regularization parameter $\lambda > 0$. In practice, such problems are typically formulated when we have access to noise-corrupted measurements $\mathbf{y} = \mathbf{y}_0 + \mathbf{n}$ where $\mathbf{n} \in \mathbb{R}^M$ is a noise term. In this case, we solve the following optimization problem

$$\mathcal{V}_\lambda \overset{\text{def.}}{=} \underset{f \in \mathrm{BV}^{(2)}(\mathbb{R})}{\arg\min} \sum_{m=1}^{M} E(f(x_m), y_m) + \lambda \|\mathrm{D}^2 f\|_{\mathcal{M}}, \tag{g-BLASSO}$$

where $E(\cdot, y)$ is a strictly convex, coercive, and differentiable cost function (typically quadratic, *i.e.*, $E(z, y) = \frac{1}{2}(z - y)^2$) for any $y \in \mathbb{R}$, and $\lambda > 0$ is a regularization parameter. The latter controls the weight between the data fidelity term $\sum_{m=1}^{M} E(f(x_m), y_m)$ and the regularization term $\|\mathrm{D}^2 f\|_{\mathcal{M}}$, and should therefore be adapted to the noise level.

5.1. From the (g-BPC) to the (g-BLASSO): Reduction to the noiseless case

We now show that the (g-BLASSO) can be reduced to an optimization problem of the form (g-BPC) (see [7, Theorem 5]), as is often done in finite-dimensional optimization problems [89, Lemma 1].

Proposition 7 (*Reformulation of the (g-BLASSO) as a (g-BPC) Problem*). *Let $\mathbf{x} \in \mathbb{R}^M$ be the ordered sampling locations, and $\mathbf{y} \in \mathbb{R}^M$ with $M \geq 2$. Let $E : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$ be a cost function such that $E(\cdot, y)$ is strictly convex, coercive, and differentiable for every $y \in \mathbb{R}$. Then, there exists a unique $\mathbf{y}_\lambda \in \mathbb{R}^M$ such that, for any $f_{\text{opt}} \in \mathcal{V}_\lambda$, $f_{\text{opt}}(x_m) = y_{\lambda,m}$ for all $m \in \{1, \dots, M\}$. Moreover, we have that the (g-BLASSO) is equivalent to the (g-BPC) with the measurement vector $\mathbf{y}_0 = \mathbf{y}_\lambda$*, i.e.,

$$\mathcal{V}_\lambda = \underset{\substack{f \in \mathrm{BV}^{(2)}(\mathbb{R}) \\ f(x_m) = y_{\lambda,m},\ m=1,\dots,M}}{\arg\min} \|\mathrm{D}^2 f\|_{\mathcal{M}}. \tag{35}$$

The proof of Proposition 7 is provided in Appendix G. The implications of this result for our problem are huge: it implies that all the results of Section 3—in particular, uniqueness, form the solutions, and sparsest solutions—can be applied to the penalized problem (g-BLASSO). The only—but crucial—catch is that the samples $\mathbf{y}_\lambda \in \mathbb{R}^M$ are unknown. Fortunately, the following proposition enables us to compute them through a standard ℓ_1-regularized discrete optimization.

Proposition 8. *Assume that the hypotheses of Proposition 7 are met. Then, the vector $\mathbf{y}_\lambda \in \mathbb{R}^M$ defined in Proposition 7 is the unique solution of the discrete minimization problem*

$$\mathbf{y}_\lambda = \underset{\mathbf{z} \in \mathbb{R}^M}{\arg\min} \sum_{m=1}^{M} E(z_m, y_m) + \lambda \|\mathbf{L}\mathbf{z}\|_1, \tag{36}$$

where $\mathbf{L} \in \mathbb{R}^{(M-2)\times M}$ is given by

$$\mathbf{L} \overset{\text{def.}}{=} \begin{pmatrix} v_1 & -(v_1+v_2) & v_2 & 0 & \cdots & 0 \\ 0 & v_2 & -(v_2+v_3) & v_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & v_{M-2} & -(v_{M-2}+v_{M-1}) & v_{M-1} \end{pmatrix}, \tag{37}$$

and $\mathbf{v} \overset{\text{def.}}{=} (v_1, \dots, v_{M-1}) \in \mathbb{R}^{M-1}$ is defined as $v_m \overset{\text{def.}}{=} \frac{1}{x_{m+1}-x_m}$ for $m \in \{1, \dots, M-1\}$.

Proof. In this proof, we denote by $f_{\mathbf{z}}$ the canonical solution (defined in Definition 5) of the (g-BPC) with sampling locations $\mathbf{x}$ and data point $\mathbf{y}_0 = \mathbf{z}$. Let us first prove that if $\mathbf{z}_{\text{opt}} \in \mathbb{R}^M$ is a solution of problem (36), then $f_{\mathbf{z}_{\text{opt}}} \in \mathrm{BV}^{(2)}(\mathbb{R})$ is a solution of the (g-BLASSO). We then deduce that for all $m \in \{1, \dots, M\}$, $z_m = f_{\mathbf{z}_{\text{opt}}}(x_m) = y_{\lambda,m}$ (where the last equality is true thanks to Proposition 7), which proves the desired result, *i.e.*, $\mathbf{y}_\lambda = \mathbf{z}_{\text{opt}}$ is the unique solution of problem (36).

Let $\mathbf{z} \in \mathbb{R}^M$. Using Eqs. (20) and (21), we have that $\|\mathrm{D}^2 f_{\mathbf{z}}\|_{\mathcal{M}} = \sum_{m=2}^{M-1} |a_m|$, where $a_m = \frac{z_{m+1}-z_m}{x_{m+1}-x_m} - \frac{z_m - z_{m-1}}{x_m - x_{m-1}}$. Therefore, we have $\|\mathrm{D}^2 f_{\mathbf{z}}\|_{\mathcal{M}} = \|\mathbf{L}\mathbf{z}\|_1$, where $\mathbf{L}$ is given by Eq. (37). This yields $\sum_{m=1}^{M} E(f_{\mathbf{z}}(x_m), y_m) + \lambda \|f_{\mathbf{z}}\|_{\mathcal{M}} = \sum_{m=1}^{M} E(z_m, y_m) + \lambda \|\mathbf{L}\mathbf{z}\|_1$. Applied to the particular case $\mathbf{z} = \mathbf{y}_\lambda$, we obtain the equality $\sum_{m=1}^{M} E(y_{\lambda,m}, y_m) + \lambda \|\mathbf{L}\mathbf{y}_\lambda\|_1 = \mathcal{J}_\lambda$, where $\mathcal{J}_\lambda$ is the optimal cost of the (g-BLASSO), since by Proposition 5, $f_{\mathbf{y}_\lambda} \in \mathcal{V}_\lambda$. This proves that the optimal value of problem (36) is lower than or equal to $\mathcal{J}_\lambda$.

Next, let $\mathbf{z}_{\text{opt}}$ be a solution of problem (36) (which exists due to the coercivity of $E(\cdot, y)$ for any $y \in \mathbb{R}$). We thus have from before that

$$\mathcal{J}_\lambda \leq \sum_{m=1}^{M} E(f_{\mathbf{z}_{\text{opt}}}(x_m), y_m) + \lambda \|\mathrm{D}^2 f_{\mathbf{z}_{\text{opt}}}\|_{\mathcal{M}} = \sum_{m=1}^{M} E(z_m, y_m) + \lambda \|\mathbf{L}\mathbf{z}_{\text{opt}}\|_1 \leq \mathcal{J}_\lambda, \tag{38}$$

which yields the desired result $f_{\mathbf{z}_{\text{opt}}} \in \mathcal{V}_\lambda$.

5.2. Algorithm for reaching a sparsest solution of the (g-BLASSO)

By combining results from the previous sections, we now formulate the following simple algorithmic pipeline to reach a sparsest solution of the (g-BLASSO).

Proposition 9. *Let $\mathbf{x} \in \mathbb{R}^M$ be the ordered sampling locations and $\mathbf{y} \in \mathbb{R}^M$ with $M \geq 2$, and let $E : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$ be a cost function such that $E(\cdot, y)$ is strictly convex, coercive, and differentiable for any $y \in \mathbb{R}$. Let the function f_{opt} be obtained through the following two-step procedure:*

1. *Compute* $\mathbf{y}_\lambda \in \mathbb{R}^M$ *(defined in Proposition 7) by solving problem (36);*
2. *Apply Algorithm 1 with the measurement vector* $\mathbf{y}_0 = \mathbf{y}_\lambda$ *to compute a sparsest solution* f_{opt} *of the (g-BPC) given by Eq. (35).*

Then, f_{opt} *is one of the sparsest solutions to the (g-BLASSO), with sparsity* $K_{min}(\mathbf{x}, \mathbf{y}_\lambda)$ *as defined in Theorem 4.*

Proof. Proposition 7 guarantees that the (g-BLASSO) is equivalent to the (g-BPC) with the measurement vector $\mathbf{y}_0 = \mathbf{y}_\lambda$. Proposition 8 then specifies that $\mathbf{y}_\lambda$ can be computed by solving problem (36). Finally, as demonstrated in the proof of Theorem 4, the output f_{opt} of Algorithm Section 4.2 reaches a sparsest solution of the corresponding (g-BPC) problem, which thus has sparsity $K_{\min}(\mathbf{x}, \mathbf{y}_\lambda)$.

Proposition 9 proposes a simple but very powerful algorithm. It reaches a sparsest solution of the (g-BLASSO)—a challenging task a priori—in two simple steps. The first consists in solving a standard ℓ_1-regularized discrete problem, for which many off-the-shelf solvers such as ADMM [90] are available. The second is our proposed sparsifying procedure, which converges in finite time. The following remarks can be made concerning Proposition 9.

Remark 2. Algorithm 1 still converges to a solution of the (g-BLASSO) when E is only a convex function, and not strictly convex as assumed in Propositions 7 and 8. The difference is that Proposition 7 no longer holds true in that there is no unique vector of measurements $\mathbf{y}_\lambda$. The solution set of the constrained problem (35) is thus in general a strict subset of $\mathcal{V}_\lambda$. Hence, the obtained solution is not necessarily the sparsest solution of the full solution set $\mathcal{V}_\lambda$, but only of this subset.

As for the assumption that E is differentiable, it is not a requirement for Proposition 9. However, as it is needed later on in Proposition 10, we include it in order to have consistent assumptions concerning E throughout the paper.

5.3. Computational complexity

The computational bottleneck of the pipeline described in Proposition 9 is its item 1; as an illustration, for $M = 50$ data points, item 1 runs in about 200 ms on commodity hardware, compared to 2 ms for item 2. This gap is due to the absence of a closed-form solution to Problem (36) owing to the non-differentiable ℓ_1 term. The latter is thus typically solved using an iterative procedure that does not converge in finite time, such as ADMM. It is well known that ADMM has a $\mathcal{O}(1/k)$ convergence rate in general, where k is the number of iterations [91]. In our case, when E is strongly convex with Lipschitz-continuous gradient, *e.g.*, with a standard quadratic loss, ADMM achieves a linear convergence rate [92]. In general, the cost per iteration of ADMM depends on how the $\mathbf{z}$-minimization step is performed, which may depend on the choice of E. In the standard quadratic case, this step consists in applying the inverse of an $M \times M$ matrix, which is fixed across iterations, to an iteration-dependent vector. To achieve this, the inverse matrix must either be computed beforehand (which is our approach), or this inverse must by applied in a matrix-free fashion. In our approach, the computational bottleneck at each iteration being the storage of the inverse matrix and its application to a vector, the computational complexity per iteration of ADMM is $\mathcal{O}(M^2)$ both in time and space.

5.4. Range of the regularization parameter λ

In practice, the choice of the regularization parameter λ is the critical element that determines the performance of our algorithm. Although this choice is highly data-dependent, in this section, we show that the search can be restricted to a bounded interval. The lower bound is $\lambda \to 0$, which corresponds at the limit to exact interpolation, that is the (g-BPC). The upper bound $\lambda \to +\infty$ corresponds to the linear regression regime, which is described in the following proposition.

Proposition 10 (*Linear Regression Regime of the (g-BLASSO)*)**.** *Let* $\mathbf{x} \in \mathbb{R}^M$ *be the ordered sampling locations and* $\mathbf{y} \in \mathbb{R}^M$ *with* $M \geq 2$*. Let* $E : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$ *be a cost function such that* $E(\cdot, y)$ *is strictly convex, coercive, and differentiable for any* $y \in \mathbb{R}$*. Then, the following properties hold.*

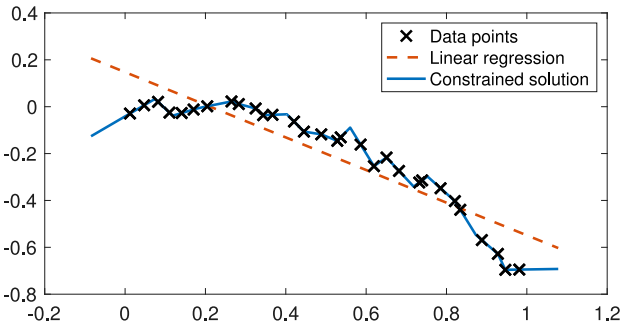
1. *There is a unique solution* $(\beta_{\text{opt}_0}, \beta_{\text{opt}_1}) \in \mathbb{R}^2$ *to the linear regression problem*

$$(\beta_{\text{opt}_0}, \beta_{\text{opt}_1}) \overset{\text{def.}}{=} \underset{(\beta_0, \beta_1) \in \mathbb{R}^2}{\arg\min} \sum_{m=1}^{M} E(\beta_0 + \beta_1 x_m, y_m). \tag{39}$$

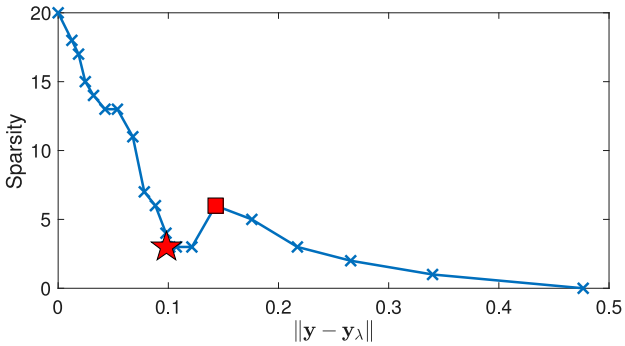
We can thus define the value

$$\lambda_{max} \overset{\text{def.}}{=} \left\| \mathbf{L}^{T\dagger} \begin{pmatrix} \partial_1 E(\beta_{\text{opt}_0} + \beta_{\text{opt}_1} x_m, y_1) \\ \vdots \\ \partial_1 E(\beta_{\text{opt}_0} + \beta_{\text{opt}_1} x_M, y_M) \end{pmatrix} \right\|_\infty, \tag{40}$$

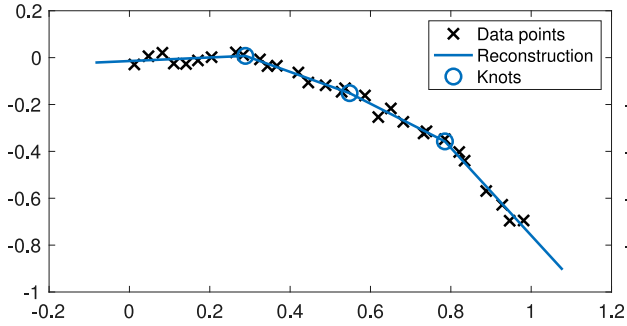
where $\partial_1 E$ *denotes the partial derivative with respect to the first variable of* E*, the matrix* $\mathbf{L}^{T\dagger}$ *denotes the pseudoinverse of* $\mathbf{L}^T$*, and* $\mathbf{L}$ *is defined as in (37).*



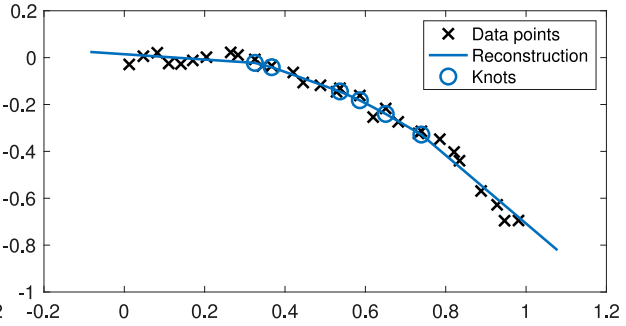
(a) Extreme cases.



(b) Sparsity *versus* error trade-off. The reconstruction corresponding to the star point is shown in Figure 7c, and the one corresponding to the square point in Figure 7d.



(c) $\lambda = 1.7 \times 10^{-3}$, loss $\|\mathbf{y} - \mathbf{y}_\lambda\| = 0.0983$, sparsity $K_{\min}(\mathbf{x}, \mathbf{y}_\lambda) = 3$.



(d) $\lambda = 1.71 \times 10^{-2}$, loss $\|\mathbf{y} - \mathbf{y}_\lambda\| = 0.1429$, sparsity $K_{\min}(\mathbf{x}, \mathbf{y}_\lambda) = 6$.

Fig. 7. Example of reconstruction for varying regularization $0 \leq \lambda \leq \lambda_{\max} = 0.1713$ with $M = 30$ simulated data points.

2. *For any $\lambda \geq \lambda_{max}$, the solution to the discrete problem (36) is given by $\mathbf{y}_\lambda = \beta_{\text{opt}_0}\mathbf{1} + \beta_{\text{opt}_1}\mathbf{x}$, where $\mathbf{1} \overset{\text{def.}}{=} (1, \dots, 1) \in \mathbb{R}^M$.*
3. *For any $\lambda \geq \lambda_{max}$, the solution to the (g-BLASSO) is unique and is the linear function f_{max} given by $f_{max}(x) \overset{\text{def.}}{=} \beta_{\text{opt}_0} + \beta_{\text{opt}_1} x$.*

The proof of Proposition 10 is given in Appendix H. Proposition 10 guarantees that the range of λ can be restricted to the interval $(0, \lambda_{\max}]$: indeed, all values $\lambda \geq \lambda_{\max}$ lead to linear regression. Moreover, the value of $\lambda_{\max}$ given in (40) only depends on the data $\mathbf{x}, \mathbf{y} \in \mathbb{R}^M$ and is easy to compute numerically—the most costly step being the computation of the pseudoinverse $\mathbf{L}^{T\dagger}$. Note that item 2 in Proposition 10, which stems from duality theory, is a generalization of a well-known result for the LASSO problem [93, Proposition 1.3], which plays a crucial role in the homotopy method [94]. The difference here is the presence of a non-invertible regularization matrix $\mathbf{L}$ in problem (36), which requires additional arguments in the proof.

6. Experiments

In this section, we describe the implementation of our two-step algorithm presented in Section 5.2 and show our experimental results. The first step of our algorithm—which consists in solving problem (36) with ADMM—is implemented using GlobalBioIm, a Matlab inverse-problem library developed by the Biomedical Imaging Group at EPFL [95]. In all our experiments, we choose the standard quadratic data fidelity loss $E(z, y) = \frac{1}{2}(z-y)^2$. This choice leads to $\partial_1 E(z, y) = z - y$, which enables the simple computation of $\lambda_{\max}$ using (40).

We present an illustrative example with $M = 30$ simulated data points in Fig. 7. A small number is chosen for visualization purposes; an application of our algorithm with a larger number of $M = 200$ data points was shown in Fig. 1. The sampling locations x_m are generated following a uniform distribution in the $[\frac{m-1}{M}, \frac{m}{M}]$ intervals for $m = 1, \dots, M$. Next, the ground-truth signal, a piecewise-linear spline f_0 in the sense of Definition 1 with 2 knots, is generated, with random knot locations τ_m within the interval $[0, 1]$, and i.i.d. Gaussian amplitudes a_m ($\sigma_a^2 = 1$). We then have $y_m = f_0(x_m) + n_m$ for $m = 1, \dots, M$, where $\mathbf{n} \in \mathbb{R}^M$ is i.i.d. Gaussian noise ($\sigma_n^2 = 4 \times 10^{-4}$).

6.1. Extreme values of λ

The reconstructions using our algorithm for extreme values of λ—*i.e.*, $\lambda \to 0$ which leads to exact interpolation of the data, and $\lambda = \lambda_{\max}$ which leads to linear regression—are shown in Fig. 7(a). Clearly, none of these solutions are satisfactory: on one hand, linear regression is too simple to model the data adequately. On the other hand, the exact interpolator suffers from overfitting. Although thanks to the sparsification procedure in Algorithm 1, its sparsity $K_{\min}(\mathbf{x}, \mathbf{y}_\lambda) = 20$ is smaller than the theoretical bound $M - 2 = 28$ given by Theorem 1, it is still clearly much larger than the desired outcome.

6.2. *Sparsity* versus *data fidelity loss trade-off*

Next, we show the sparsity $K_{\min}(\mathbf{x}, \mathbf{y}_\lambda)$ *versus* error $\|\mathbf{y} - \mathbf{y}_\lambda\|$ trade-off curve in Fig. 7(b). The latter was obtained by applying our algorithm with 20 values of λ (equispaced on a logarithmic scale) within the range $[\lambda_{\min}, \lambda_{\max}]$, with $\lambda_{\max} = 0.1713$ (as defined in (40)) and $\lambda_{\min} \overset{\text{def.}}{=} 10^{-5} \times \lambda_{\max}$. We thus observe the evolution from exact interpolation to linear regression as λ increases.

Ideally, one would like to choose to value of λ that minimizes $\|\mathbf{y}_0 - \mathbf{y}_\lambda\|$, *i.e.*, the error with respect to the noiseless data $\mathbf{y}_0$. However, in practice, the noiseless data is unknown, and one must use the noisy data $\mathbf{y}$. Depending on the noise level, solely minimizing $\|\mathbf{y} - \mathbf{y}_\lambda\|$ might not be a desirable objective, since it leads to overfitting. Hence, we consider the trade-off between data fidelity loss and sparsity as a proxy for the standard universality *versus* simplicity trade-off in machine learning. Note that we choose the data fidelity loss $\|\mathbf{y} - \mathbf{y}_\lambda\|$ instead of λ as the x-axis metric, since it is an increasing function of the latter, and the former is easier to interpret.

This trade-off curve does not specify a single optimal value of the regularization parameter λ. Instead, it helps the user choose an appropriate balance by giving quantitative, interpretable data about the possible trade-offs. A key observation is that this curve is not necessarily monotonous: the sparsity can increase as $\|\mathbf{y} - \mathbf{y}_\lambda\|$ increases, as shown in Fig. 7(b). This lack of monotonicity is rather counter-intuitive, since the overall trend as λ increases is to go from sparsity $K_{\min}(\mathbf{x}, \mathbf{y}) = 20$ to $K_{\min}(\mathbf{x}, \mathbf{y}_{\lambda_{\max}}) = 0$. Note that a similar behavior has been known to occur in the context of the homotopy method [93], although it is far from being systematic. However, the interesting feature is that, in the sparsity *versus* error trade-off, some values of λ are sometimes strictly better than others for both metrics, such as the star point over the square point in Fig. 7(b). Having access to the full trade-off curve such as Fig. 7(b) is very helpful to judiciously select a suitable value of λ. This holds true as well when the curve is monotonic: indeed, the user should select the value of λ such that the data fidelity is lowest for the desired level of sparsity, *i.e.*, the leftmost point of every plateau.

6.3. Example reconstructions

To illustrate the non-monotonicity of the sparsity *versus* error curve, examples of reconstructions for two specific values of λ are shown in Figs. 7(c) and 7(d). Indeed, the former reconstruction has a lower value of λ, and thus lower data-fidelity loss. Nevertheless, the reconstruction in Fig. 7(c) is sparser, with $K_{\min}(\mathbf{x}, \mathbf{y}_\lambda) = 3$ *versus* 6 in Fig. 7(d). Note that this gap is not a numerical artefact, since the magnitude of the weights $\tilde{a}_k$ associated to the knots in Fig. 7(d) is much greater than numerical precision. This indicates that the value of λ for Fig. 7(c) should be preferred to that of 7(d).

7. Conclusion

In this paper, we fully described the solution set of the (g-BPC), which consists in interpolating data points by minimizing the TV norm of the second derivative. More precisely, we specified the cases in which it has a unique solution, the form of all the solutions, and the subset of sparsest solutions. We also proposed a simple and fast algorithm to reach (one of) the sparsest solution(s). We then extended these results to the (g-BLASSO), by showing that it can be reformulated as a (g-BPC) problem. Next, we introduced a two-step algorithm to solve the (g-BLASSO), the first step of which consists in solving a discrete ℓ_1-regularized problem, and the second in applying our algorithm to solve a (g-BPC) problem. Finally, we applied our algorithm to some simulated data, and suggested plotting the sparsity *versus* data fidelity error plot in order to judiciously select a suitable value of the regularization parameter. This paper paves the way for the study of supervised learning problems through the formulation of variational inverse problems with TV-based regularization, by completely describing the one-dimensional scenario. A future exciting—albeit much more challenging—prospect would be to achieve similar results in higher dimensions, *i.e.*, to reconstruct functions $f : \mathbb{R}^d \to \mathbb{R}$ with $d > 1$. This would be a major milestone to better understand ReLU networks and deep learning in general, whose practical outstanding performances are yet to be fully explained.

Acknowledgments

The authors are thankful to Shayan Aziznejad for many discussions related to this work and for his elegant connection between the (g-BLASSO) problem and its discrete counterpart (see (36)). Julien Fageot was supported by the Swiss National Science Foundation (SNSF) under Grants P2ELP2_181759 and P400P2_194364. The work of Thomas Debarre, Quentin Denoyelle, and Michael Unser is supported by the SNSF under Grant 200020_184646/1 and the European Research Council (ERC) under Grant 692726-GlobalBioIm.

Appendix A. The space $\mathrm{BV}^{(2)}(\mathbb{R})$

As a complement to the characterization of the space $\mathrm{BV}^{(2)}(\mathbb{R})$ in Section 2.2, we summarize its main properties in Proposition 11, revealing its Banach-space structure. The construction of the native space for general spline-admissible operator L (we consider here the case $\mathrm{L} = \mathrm{D}^2$) is developed in [69].

Proposition 11 (*Properties of* $\mathrm{BV}^{(2)}(\mathbb{R})$)**.** *The space* $\mathrm{BV}^{(2)}(\mathbb{R})$ *has the following properties.*

1. *Any function* $f \in \mathrm{BV}^{(2)}(\mathbb{R})$ *is continuous and satisfies* $f(x) = \mathcal{O}(x)$ *at infinity. Affine functions* f *such that* $f(x) = ax + b$ *for* $a, b \in \mathbb{R}$ *are elements of* $\mathrm{BV}^{(2)}(\mathbb{R})$.
2. *The linear space* $\mathrm{BV}^{(2)}(\mathbb{R})$ *is isomorphic to* $\mathcal{M}(\mathbb{R}) \times \mathbb{R}^2$ *via the relation*
$$f \mapsto \left(\mathrm{D}^2 f, (f(0), f(1) - f(0))\right). \tag{A.1}$$
3. *The space* $\mathrm{BV}^{(2)}(\mathbb{R})$ *is a Banach space for the norm*
$$\|f\|_{\mathrm{BV}^{(2)}} \overset{\text{def.}}{=} \|\mathrm{D}^2 f\|_{\mathcal{M}} + \sqrt{f(0)^2 + (f(1) - f(0))^2}. \tag{A.2}$$
4. *For any* $w \in \mathcal{M}(\mathbb{R})$*, there exists a unique* $f \in \mathrm{BV}^{(2)}(\mathbb{R})$ *such that* $\mathrm{D}^2 f = w$ *and* $f(0) = f(1) = 0$.

Proof. A function in $\mathrm{BV}^{(2)}(\mathbb{R})$ is the integration of a bounded-variation function, and is therefore continuous. If f is such that $\mathrm{D}^2 f \in \mathcal{M}(\mathbb{R})$, then $\mathrm{D}f$ is bounded by $\|\mathrm{D}^2 f\|_{\mathcal{M}}$. Hence,
$$|f(x)| = \left| f(0) + \int_0^x (\mathrm{D}f)(t)\mathrm{d}t \right| \leq |f(0)| + \|\mathrm{D}f\|_\infty |x|, \tag{A.3}$$
and $f(x) = \mathcal{O}(x)$ at infinity. Moreover, for an affine function f such that $f(x) = a + bx$, we obviously have that $\mathrm{D}^2 f = 0 \in \mathcal{M}(\mathbb{R})$, hence $f \in \mathrm{BV}^{(2)}(\mathbb{R})$. The relation (A.1) is clearly linear and is a bijection, since any $f \in \mathrm{BV}^{(2)}(\mathbb{R})$ can be uniquely recovered from its second derivative via the specification of two boundary conditions, here the values of $f(0)$ and $f(1)$. Hence, (A.1) is an isomorphism.

Due to this isomorphism, $\mathrm{BV}^{(2)}(\mathbb{R})$ inherits the Banach space structure of $\mathcal{M}(\mathbb{R}) \times \mathbb{R}^2$ for the norm $\|(w, (\beta_0, \beta_1))\|_{\mathcal{M}\times\mathbb{R}^2} = \|w\| + \sqrt{\beta_0^2 + \beta_1^2}$ and is hence a Banach space for the norm (A.2). For the last point, by definition, any $f \in \mathcal{S}'(\mathbb{R})$ such that $\mathrm{D}^2 f = w$ is in $\mathrm{BV}^{(2)}(\mathbb{R})$. The space of solutions of $\mathrm{D}^2 f = w$ is then a two-dimensional space, and the solution is uniquely characterized by the specification of the two boundary conditions $f(0) = 0$ and $f(1) = 0$.

In Section 2.2, we have introduced the operator D_0^{-2}. We now summarize its main properties.

Proposition 12 (*Kernel of* D_0^{-2})**.** *For any* $w \in \mathcal{M}(\mathbb{R})$, $\mathrm{D}_0^{-2}\{w\}$ *is given by*
$$\mathrm{D}_0^{-2}\{w\}(x) \overset{\text{def.}}{=} \int_{\mathbb{R}} g(x, y)\mathrm{d}w(y) = \langle w, g(x, \cdot)\rangle, \tag{A.4}$$
where g *is the kernel defined over* $\mathbb{R}^2$ *as*
$$g(x, y) \overset{\text{def.}}{=} (x - y)_+ - (-y)_+ + x\left((-y)_+ - (1 - y)_+\right), \tag{A.5}$$
and is such that $g(x, \cdot)$ *is a continuous and compactly supported function for any* $x \in \mathbb{R}$*. Then, the operator* D_0^{-2} *is linear and continuous from* $\mathcal{M}(\mathbb{R})$ *to* $\mathrm{BV}^{(2)}(\mathbb{R})$ *and satisfies the right-inverse and pseudo-left-inverse relations*
$$\forall w \in \mathcal{M}(\mathbb{R}), \quad \mathrm{D}^2\{\mathrm{D}_0^{-2}\{w\}\} = w, \tag{A.6}$$
$$\forall f \in \mathrm{BV}^{(2)}(\mathbb{R}), \ \forall x \in \mathbb{R}, \quad f(x) = \mathrm{D}_0^{-2}\{\mathrm{D}^2\{f\}\}(x) + f(0) + (f(1) - f(0))x. \tag{A.7}$$
In particular, D_0^{-2} *is a right-inverse of the second-derivative* D^2*. Moreover, any* $f \in \mathrm{BV}^{(2)}(\mathbb{R})$ *can be uniquely decomposed as*
$$\forall x \in \mathbb{R}, \quad f(x) = \mathrm{D}_0^{-2}\{w\}(x) + \beta_0 + \beta_1 x, \tag{A.8}$$
where $w \in \mathcal{M}(\mathbb{R})$, $\beta_0, \beta_1 \in \mathbb{R}$ *are given by*
$$w = \mathrm{D}^2 f, \quad \beta_0 = f(0), \quad \text{and} \quad \beta_1 = f(1) - f(0). \tag{A.9}$$

Proof. We fix $x \in \mathbb{R}$. We easily verify that $g(x, y) = 0$ for $|y| \geq \max(1, |x|)$, hence $g(x, \cdot)$ is compactly supported. The function $g(x, \cdot)$ is continuous due to the continuity of $y \mapsto y_+$. Therefore, $g(x, \cdot) \in \mathcal{C}_0(\mathbb{R})$ and the duality product $\langle w, g(x, \cdot)\rangle$ is well defined for any $w \in \mathcal{M}(\mathbb{R})$ and $x \in \mathbb{R}$.

For $w \in \mathcal{M}(\mathbb{R})$ and $x \in \mathbb{R}$, we set $f(x) = \langle w, g(x, \cdot)\rangle$. We now prove that $\mathrm{D}^2 f = w$ in the distributional sense. First, we prove that f is continuous and is therefore an element of the space of distributions $\mathcal{D}'(\mathbb{R})$. For any $x, x_0 \in \mathbb{R}$, we have that $|f(x) - f(x_0)| = \left|\int_{\mathbb{R}} (g(x, y) - g(x_0, y))\mathrm{d}w(y)\right| \leq \|g(x, \cdot) - g(x_0, \cdot)\|_\infty \|w\|_{\mathcal{M}}$, and we easily see from the definition of

g in (A.5) that $\|g(x,\cdot)-g(x_0,\cdot)\|_\infty \to 0$ when $x \to x_0$. It then suffices to show that $\langle \mathrm{D}^2 f, \varphi\rangle = \langle w, \varphi\rangle$ for any compactly supported and infinitely smooth test function $\varphi \in \mathcal{D}(\mathbb{R})$ to deduce that $\mathrm{D}^2 f = w$ in $\mathcal{D}'(\mathbb{R})$, and that this equality also holds in $\mathcal{M}(\mathbb{R})$ since $w \in \mathcal{M}(\mathbb{R})$.

From the definition of g, denoting by ∂_x the partial derivative with respect to the first variable, we have

$$\partial_x^2\{g\}(\cdot,y) = \delta(\cdot - y). \tag{A.10}$$

Let $\varphi \in \mathcal{D}(\mathbb{R})$ and K be its compact support. We have that

$$\int_{\mathbb{R}}\int_{\mathbb{R}} |g(x,y)||\varphi''(x)|\mathrm{d}w(y)\mathrm{d}x \le \|\varphi''\|_\infty \sup_{x\in K,\ y\in\mathbb{R}} |g(x,y)|\mathrm{Leb}(K)\,\|w\|_{\mathcal{M}}\,, \tag{A.11}$$

where $\mathrm{Leb}(K)$ is the Lebesgue measure of K. We then observe that, for any fixed $x \in \mathbb{R}$, we have $|g(x,y)| \le |(x-y)_+ - (-y)_+| + |x||(1-y)_+ - (-y)_+| \le |x| + |x| \le 2|x|$, hence $\sup_{x\in K,\ y\in\mathbb{R}} |g(x,y)| \le 2\sup_{x\in K}|x| < \infty$ and therefore

$$\int_{\mathbb{R}}\int_{\mathbb{R}} |g(x,y)||\varphi''(x)|\mathrm{d}w(y)\mathrm{d}x < \infty. \tag{A.12}$$

Then, we have that

$$\langle \mathrm{D}^2 f, \varphi\rangle = \langle f, \mathrm{D}^2\varphi\rangle = \int_{\mathbb{R}}\left(\int_{\mathbb{R}} g(x,y)\mathrm{d}w(y)\right)\varphi''(x)\mathrm{d}x = \int_{\mathbb{R}}\left(\int_{\mathbb{R}} \varphi''(x)g(x,y)\mathrm{d}x\right)\mathrm{d}w(y), \tag{A.13}$$

where the second equality follows from Fubini's theorem with the hypothesis (A.12).

Due to (A.10), we then observe that $\int_{\mathbb{R}} \varphi''(x)g(x,y)\mathrm{d}x = \langle \mathrm{D}^2\varphi, g(\cdot,y)\rangle = \langle \varphi, \partial_x^2 g(\cdot,y)\rangle = \langle \varphi, \delta(\cdot-y)\rangle = \varphi(y)$. Hence, (A.13) yields $\langle \mathrm{D}^2 f, \varphi\rangle = \int_{\mathbb{R}} \varphi(y)\mathrm{d}w(y) = \langle w, \varphi\rangle$, which proves that $\mathrm{D}^2 f = w$.

Moreover, we have that $g(0,y) = g(1,y) = 0$ for all $y \in \mathbb{R}$, which yields $f(0) = f(1) = 0$. From the definition of D_0^{-2}, $\mathrm{D}_0^{-2}\{w\}$ is the unique function satisfying these properties, proving that $\mathrm{D}_0^{-2}\{w\}(x) = f(x) = \langle w, g(x,\cdot)\rangle$ for every $x \in \mathbb{R}$ and $w \in \mathcal{M}(\mathbb{R})$. This shows (A.4).

Next, it is clear that D_0^{-2} is linear from $\mathcal{M}(\mathbb{R})$ to $\mathrm{BV}^{(2)}(\mathbb{R})$. The continuity of D_0^{-2} follows from the fact that

$$\|\mathrm{D}_0^{-2}\{w\}\|_{\mathrm{BV}^{(2)}} = \|\mathrm{D}^2\mathrm{D}_0^{-2}\{w\}\|_{\mathcal{M}} + \sqrt{(\mathrm{D}_0^{-2}\{w\}(0))^2 + (\mathrm{D}_0^{-2}\{w\}(1) - \mathrm{D}_0^{-2}\{w\}(0))^2} = \|w\|_{\mathcal{M}}. \tag{A.14}$$

The equality $\mathrm{D}^2\mathrm{D}_0^{-2}\{w\} = w$ comes from the definition of $\mathrm{D}_0^{-2}\{w\}$. For the right-hand side of (A.7), we remark that $\mathrm{D}^2\{\mathrm{D}_0^{-2}\mathrm{D}^2\{f\}\} = \mathrm{D}^2 f$ by definition, hence $\mathrm{D}_0^{-2}\mathrm{D}^2\{f\}(x) = f(x) + \beta_0 + \beta_1 x$ for every $x \in \mathbb{R}$ and some constants $\beta_0, \beta_1 \in \mathbb{R}$. The equations $\mathrm{D}_0^{-2}\mathrm{D}^2\{f\}(0) = \mathrm{D}_0^{-2}\mathrm{D}^2\{f\}(1) = 0$ then specify the constants β_0 and β_1, which proves (A.7). Finally, (A.8) and (A.9) can be seen as reformulations of the right equality in (A.7). The uniqueness follows from the simple fact that $\mathrm{D}^2 f = w$ determines f when the values of $f(0)$ and $f(1)$ are fixed.

Appendix B. Proof of Proposition 1

The forward operator considered in this paper is a sampling operator (the functions $f \in \mathrm{BV}^{(2)}(\mathbb{R})$ are sampled at the locations $x_m \in \mathbb{R}$ for $m \in \{1,\dots,M\}$). Let us denote it, for the convenience of the proof, as a linear operator $\boldsymbol{\nu} : \mathrm{BV}^{(2)}(\mathbb{R}) \to \mathbb{R}^M$ such that

$$\forall f \in \mathrm{BV}^{(2)}(\mathbb{R}), \quad \boldsymbol{\nu}(f) \overset{\text{def.}}{=} (f(x_m))_{1\le m\le M}. \tag{B.1}$$

The proof of Proposition 1 can be divided in several steps. First, we reformulate (g-BPC) into an equivalent optimization problem thanks to the decomposition of any $f \in \mathrm{BV}^{(2)}(\mathbb{R})$ given by (7). This is stated in the next lemma.

Lemma 2. *The problem (g-BPC) is equivalent to*

$$\min_{(w,(\beta_0,\beta_1))\in\mathcal{M}(\mathbb{R})\times\mathbb{R}^2} \iota_{\{\mathbf{y}_0\}}(\boldsymbol{\nu}_{\mathcal{M}}(w) + \beta_0\mathbf{1} + \beta_1\mathbf{x}) + \|w\|_{\mathcal{M}}\,. \tag{B.2}$$

where $\iota_{\{\mathbf{y}_0\}}$ is the indicator of the convex set $\{\mathbf{y}_0\}$, which is zero at $\mathbf{y}_0$ and $+\infty$ elsewhere, and

$$\boldsymbol{\nu}_{\mathcal{M}} \overset{\text{def.}}{=} \boldsymbol{\nu}\circ\mathrm{D}_0^{-2} : \mathcal{M}(\mathbb{R}) \to \mathbb{R}^M \tag{B.3}$$

is the modified forward operator. This equivalence is in the sense that there exists a bijection given by the unique decomposition of any $f \in \mathrm{BV}^{(2)}(\mathbb{R})$ as $f = \mathrm{D}_0^{-2}\{w\} + \beta_0 + \beta_1(\cdot)$ with $(w,(\beta_0,\beta_1)) \in \mathcal{M}(\mathbb{R})\times\mathbb{R}^2$ (see(7)) between the solution sets of both optimization problems.

From now on, we consider the equivalent problem (B.2) and analyze it using tools from duality theory. The search space $\mathcal{M}(\mathbb{R})\times\mathbb{R}^2$ of this optimization problem is endowed with the weak-* topology, which is defined in terms of its predual space $\mathcal{C}_0(\mathbb{R})\times\mathbb{R}^2$. Using (A.4), the modified operator $\boldsymbol{\nu}_{\mathcal{M}}$ can be expressed as $\boldsymbol{\nu}_{\mathcal{M}}(w) = (\langle w, g(x_m,\cdot)\rangle)_{1\le m\le M}$, where $g(x_m,\cdot) \in \mathcal{C}_0(\mathbb{R})$ for all $m \in \{1,\dots,M\}$ by Proposition 12. Since $\mathcal{M}(\mathbb{R})$ is the dual of $\mathcal{C}_0(\mathbb{R})$, this implies that the

linear functional $\boldsymbol{\nu}_{\mathcal{M}} : \mathcal{M}(\mathbb{R}) \to \mathbb{R}^M$ is weak-* continuous [96, Theorem IV.20, p. 114]. The adjoint $\boldsymbol{\nu}_{\mathcal{M}}^* : \mathbb{R}^M \to \mathcal{C}_0(\mathbb{R})$ of $\boldsymbol{\nu}_{\mathcal{M}}$ is thus uniquely defined and is given by

$$\forall \mathbf{c} \in \mathbb{R}^M, \quad \boldsymbol{\nu}_{\mathcal{M}}^*(\mathbf{c}) = \sum_{m=1}^{M} c_m g(x_m, \cdot), \tag{B.4}$$

since $\left\langle w,\ \boldsymbol{\nu}_{\mathcal{M}}^*(\mathbf{c}) \right\rangle = \langle \boldsymbol{\nu}_{\mathcal{M}}(w),\ \mathbf{c} \rangle = \left\langle (\langle w,\ g(x_m, \cdot) \rangle)_{1 \leq m \leq M},\ \mathbf{c} \right\rangle = \left\langle w,\ \sum_{m=1}^{M} c_m g(x_m, \cdot) \right\rangle$, for all $w \in \mathcal{M}(\mathbb{R})$ and $\mathbf{c} \in \mathbb{R}^M$.

The second part of the proof consists in determining the dual problem of (B.2), proving that strong duality between the primal and dual problem holds (*i.e.*, that the optimal values of both problems are equal and finite) and then deriving the optimality conditions which characterize the solutions of problem (B.2). This is done in the next lemma.

Lemma 3. *The dual problem of (B.2) is given by*

$$\sup_{\mathbf{c} \in \mathcal{C}} \langle \mathbf{y}_0,\ \mathbf{c} \rangle, \quad \textit{with} \quad \mathcal{C} \overset{\text{def.}}{=} \{ \mathbf{c} \in \mathbb{R}^M :\ \langle \mathbf{c},\ \mathbf{1} \rangle = \langle \mathbf{c},\ \mathbf{x} \rangle = 0,\ \left\| \boldsymbol{\nu}_{\mathcal{M}}^*(\mathbf{c}) \right\|_\infty \leq 1 \}. \tag{B.5}$$

Moreover, it has at least one solution and strong duality holds between problems (B.2) and (B.5). Finally, for any $(w, (\beta_0, \beta_1)) \in \mathcal{M}(\mathbb{R}) \times \mathbb{R}^2$ and $\mathbf{c} \in \mathbb{R}^M$, we have the equivalence between the following statements:

1. *$(w, (\beta_0, \beta_1))$ is a solution of (B.2) and $\mathbf{c}$ is a solution of (B.5).*
2. *$(w, (\beta_0, \beta_1))$ and $\mathbf{c}$ satisfy the following conditions:*

$$\boldsymbol{\nu}_{\mathcal{M}}(w) + \beta_0 \mathbf{1} + \beta_1 \mathbf{x} = \mathbf{y}_0, \tag{B.6}$$

$$\langle \mathbf{c},\ \mathbf{1} \rangle = \langle \mathbf{c},\ \mathbf{x} \rangle = 0, \quad \|w\|_{\mathcal{M}} = \left\langle w,\ \boldsymbol{\nu}_{\mathcal{M}}^*(\mathbf{c}) \right\rangle \quad \textit{and} \quad \left\| \boldsymbol{\nu}_{\mathcal{M}}^*(\mathbf{c}) \right\|_\infty \leq 1. \tag{B.7}$$

Proof. Let us first obtain the dual problem (B.5). The proof follows the technique of perturbed problems detailed in [97, Chapter 3].

Dual problem. Let us write the (primal) problem (B.2) as

$$\min_{(w, (\beta_0, \beta_1)) \in \mathcal{M}(\mathbb{R}) \times \mathbb{R}^2} F(w, (\beta_0, \beta_1)) + G(\Lambda(w, (\beta_0, \beta_1))), \tag{B.8}$$

where $F(w, (\beta_0, \beta_1)) \overset{\text{def.}}{=} \|w\|_{\mathcal{M}}$, $G(\mathbf{c}) \overset{\text{def.}}{=} \iota_{\{\mathbf{y}_0\}}(\mathbf{c})$ for all $\mathbf{c} \in \mathbb{R}^M$, and $\Lambda(w, (\beta_0, \beta_1)) \overset{\text{def.}}{=} \boldsymbol{\nu}_{\mathcal{M}}(w) + \beta_0 \mathbf{1} + \beta_1 \mathbf{x}$.

The functions F and G are convex, lower semi-continuous and not identically equal to $\pm\infty$. By [97, Equation (4.18)], the dual problem of (B.8) is thus given by $\sup_{\mathbf{c} \in \mathbb{R}^M} -F^*(\Lambda^*(\mathbf{c})) - G^*(-\mathbf{c})$, where F^* and G^* are the Fenchel conjugates of F and G respectively, and $\Lambda^* : \mathbb{R}^M \to \mathcal{C}_0(\mathbb{R}) \times \mathbb{R}^2$ is the adjoint of Λ. One can check that for all $\mathbf{c} \in \mathbb{R}^M$, $G^*(\mathbf{c}) = \langle \mathbf{c},\ \mathbf{y}_0 \rangle$, for all $\eta \in \mathcal{C}_0(\mathbb{R})$ and $\beta_0, \beta_1 \in \mathbb{R}$, $F^*(\eta, (\beta_0, \beta_1)) = \iota_{\|\cdot\|_\infty \leq 1}(\eta) + \iota_{\{(0,0)\}}((\beta_0, \beta_1))$ (with $\iota_{\|\cdot\|_\infty \leq 1}$ the indicator function of the closed unit ball in $\mathcal{C}_0(\mathbb{R})$ for the uniform norm), and for all $\mathbf{c} \in \mathbb{R}^M$, $\Lambda^*(\mathbf{c}) = \left(\boldsymbol{\nu}_{\mathcal{M}}^*(\mathbf{c}), (\langle \mathbf{c},\ \mathbf{1} \rangle, \langle \mathbf{c},\ \mathbf{x} \rangle) \right)$. Therefore, the dual problem can be rewritten as

$$-\inf_{\mathbf{c} \in \mathbb{R}^M} \iota_{\mathcal{C}}(\mathbf{c}) + \langle -\mathbf{c},\ \mathbf{y}_0 \rangle, \tag{B.9}$$

where $\mathcal{C} \subset \mathbb{R}^M$ is the convex set defined in (B.5). Problem (B.9) is clearly the same as problem (B.5), which proves the first statement of the lemma.

Strong duality. To prove strong duality between problems (B.2) and (B.5) (*i.e.*, they have the same optimal value), we start by showing strong duality between

$$\inf_{\mathbf{c} \in \mathbb{R}^M} \iota_{\mathcal{C}}(\mathbf{c}) + \langle -\mathbf{c},\ \mathbf{y}_0 \rangle, \tag{B.10}$$

and its dual problem. We then conclude by observing that the optimal value of the dual problem of (B.10) is equal to the optimal value of problem (B.2) up to a sign. Indeed, this last statement proves that both problems (B.2) and (B.5) have the same optimal value since problem (B.10) is, up to a sign, the dual problem (B.5) (which rewrites as in (B.9)).

We first start by proving that strong duality holds between problem (B.10) and its dual problem. The aim is to apply [97, Proposition 2.3, Chapter 3]. With the notations of [97], let us denote the map $\Phi : \mathbb{R}^M \times \mathcal{C}_0(\mathbb{R}) \to \mathbb{R} \cup \{+\infty\}$ as

$$\forall (\mathbf{c}, \eta) \in \mathbb{R}^M \times \mathcal{C}_0(\mathbb{R}), \quad \Phi(\mathbf{c}, \eta) \overset{\text{def.}}{=} \langle -\mathbf{c},\ \mathbf{y}_0 \rangle + \iota_{\{(0,0)\}}(\langle \mathbf{c},\ \mathbf{1} \rangle, \langle \mathbf{c},\ \mathbf{x} \rangle)) + \iota_{\|\cdot\|_\infty \leq 1}(\boldsymbol{\nu}_{\mathcal{M}}^*(\mathbf{c}) - \eta). \tag{B.11}$$

This map Φ defines a perturbed problem to problem (B.10), since by definition, for all $\mathbf{c} \in \mathbb{R}^M$,

$$\Phi(\mathbf{c}, 0) = \iota_{\mathcal{C}}(\mathbf{c}) + \langle -\mathbf{c},\ \mathbf{y}_0 \rangle \tag{B.12}$$

is the objective function of problem (B.10). Now let us check that the assumptions of [97, Proposition 2.3] are satisfied for Φ and problem (B.10):

- Φ is convex,
- the optimal value of problem (B.10) is finite due to the weak duality (primal–dual inequality given below) between problems (B.8) and (B.9), which yields

$$-\infty < -\inf_{\mathbf{c}\in\mathbb{R}^M} \iota_{\mathcal{C}}(\mathbf{c}) + \langle -\mathbf{c}, \mathbf{y}_0\rangle \leq \inf_{(w,(\beta_0,\beta_1))\in\mathcal{M}(\mathbb{R})\times\mathbb{R}^2} \|w\|_{\mathcal{M}} + \iota_{\{\mathbf{y}_0\}}(\boldsymbol{\nu}_{\mathcal{M}}(w) + \beta_0\mathbf{1} + \beta_1\mathbf{x}) < +\infty, \tag{B.13}$$

- the map $\eta \in \mathcal{C}_0(\mathbb{R}) \mapsto \Phi(\mathbf{0}, \eta) = \iota_{\|\cdot\|_\infty\leq 1}(-\eta)$ is finite and continuous at $\eta = 0 \in \mathcal{C}_0(\mathbb{R})$.

Therefore, we deduce that strong duality holds between problem (B.10) and its dual problem given by

$$\sup_{w\in\mathcal{M}(\mathbb{R})} \; -\Phi^*(\mathbf{0}, w), \tag{B.14}$$

and that this last optimization problem has at least one solution. By writing the map Φ as $\Phi(\mathbf{c}, \eta) = \tilde{F}(\mathbf{c}) + \tilde{G}(\tilde{\Lambda}(\mathbf{c}) - \eta)$ with $\tilde{F}(\mathbf{c}) \overset{\text{def.}}{=} \langle -\mathbf{c}, \mathbf{y}_0\rangle + \iota_{V^\perp}(\mathbf{c})$, $V \overset{\text{def.}}{=} \text{Span}(\mathbf{1}, \mathbf{x}) \subset \mathbb{R}^M$, $\tilde{G} \overset{\text{def.}}{=} \iota_{\|\cdot\|_\infty\leq 1}(\cdot)$, and $\tilde{\Lambda} = \boldsymbol{\nu}^*_{\mathcal{M}}$,we get that $\Phi^*(\mathbf{c}, w) = \tilde{F}^*(\tilde{\Lambda}^*(w) + \mathbf{c}) + \tilde{G}^*(-w)$ for any $(\mathbf{c}, w) \in \mathbb{R}^M \times \mathcal{M}(\mathbb{R})$, and thus that problem (B.14) becomes

$$-\min_{w\in\mathcal{M}(\mathbb{R})} \iota_V(\boldsymbol{\nu}_{\mathcal{M}}(w) + \mathbf{y}_0) + \|w\|_{\mathcal{M}}\,. \tag{B.15}$$

We now verify that the optimal value of

$$\min_{w\in\mathcal{M}(\mathbb{R})} \iota_V(\boldsymbol{\nu}_{\mathcal{M}}(w) + \mathbf{y}_0) + \|w\|_{\mathcal{M}}\,, \tag{B.16}$$

i.e., minus the optimal value of the dual problem of (B.10) is equal to the optimal value of problem (B.2)

$$\min_{(w,(\beta_0,\beta_1))\in\mathcal{M}(\mathbb{R})\times\mathbb{R}^2} \iota_{\{\mathbf{y}_0\}}(\boldsymbol{\nu}_{\mathcal{M}}(w) + \beta_0\mathbf{1} + \beta_1\mathbf{x}) + \|w\|_{\mathcal{M}}\,. \tag{B.17}$$

Let $w \in \mathcal{M}(\mathbb{R})$ be a solution of problem (B.16) (which we know to exist by [97, Proposition 2.3]). Since the objective function of problem (B.16) is finite at w, we obtain that $\boldsymbol{\nu}_{\mathcal{M}}(w) + \mathbf{y}_0 \in V$, *i.e.*, there exists $(\beta_0, \beta_1) \in \mathbb{R}^2$ such that $\mathbf{y}_0 = \boldsymbol{\nu}_{\mathcal{M}}(-w) + \beta_0\mathbf{1} + \beta_1\mathbf{x}$. Assume by contradiction that there exist $(\tilde{w}, (\tilde{\beta}_0, \tilde{\beta}_1)) \in \mathcal{M}(\mathbb{R}) \times \mathbb{R}^2$ that achieve a lower cost than $(w, (\beta_0, \beta_1))$ in (B.2), *i.e.*,

$$\iota_{\{\mathbf{y}_0\}}(\boldsymbol{\nu}_{\mathcal{M}}(-w) + \beta_0\mathbf{1} + \beta_1\mathbf{x}) + \|-w\|_{\mathcal{M}} > \iota_{\{\mathbf{y}_0\}}(\boldsymbol{\nu}_{\mathcal{M}}(\tilde{w}) + \tilde{\beta}_0\mathbf{1} + \tilde{\beta}_1\mathbf{x}) + \|\tilde{w}\|_{\mathcal{M}}\,. \tag{B.18}$$

Since the left term of this inequality in finite, we must have $\mathbf{y}_0 = \boldsymbol{\nu}_{\mathcal{M}}(\tilde{w}) + \tilde{\beta}_0\mathbf{1} + \tilde{\beta}_1\mathbf{x}$ and

$$\|w\|_{\mathcal{M}} > \|-\tilde{w}\|_{\mathcal{M}}\,. \tag{B.19}$$

Since $\boldsymbol{\nu}_{\mathcal{M}}(-\tilde{w}) + \mathbf{y}_0 = \tilde{\beta}_0\mathbf{1} + \tilde{\beta}_1\mathbf{x} \in V$, we deduce thanks to (B.19) that $-\tilde{w}$ achieves a lower cost than w for problem (B.16), which contradicts the assumption on w. Hence, for all $w \in \mathcal{M}(\mathbb{R})$, $(\beta_0, \beta_1) \in \mathbb{R}^2$, we have

$$\iota_{\{\mathbf{y}_0\}}(\boldsymbol{\nu}_{\mathcal{M}}(-w) + \beta_0\mathbf{1} + \beta_1\mathbf{x}) + \|-w\|_{\mathcal{M}} \leq \iota_{\{\mathbf{y}_0\}}(\boldsymbol{\nu}_{\mathcal{M}}(w) + \beta_0\mathbf{1} + \beta_1\mathbf{x}) + \|w\|_{\mathcal{M}}\,, \tag{B.20}$$

i.e., $(-w, (\beta_0, \beta_1)) \in \mathcal{M}(\mathbb{R}) \times \mathbb{R}^2$ is a solution of problem (B.2). Therefore, we get that the optimal values of problems (B.2) and (B.16) are equal since

$$\iota_V(\boldsymbol{\nu}_{\mathcal{M}}(w) + \mathbf{y}_0) + \|w\|_{\mathcal{M}} = \iota_{\{\mathbf{y}_0\}}(\boldsymbol{\nu}_{\mathcal{M}}(-w) + \beta_0\mathbf{1} + \beta_1\mathbf{x}) + \|-w\|_{\mathcal{M}}\,. \tag{B.21}$$

Optimality conditions. To derive the optimality conditions given in (B.6) and (B.7), we apply [97, Proposition 2.4, Chapter 3]. We have already proved that strong duality holds, and that the primal problem (B.2) has at least one solution. To apply the proposition, it remains to prove that the dual problem (B.5) also has at least one solution. This holds true due to the following

- the objective function of problem (B.5) is a continuous linear form over the convex set $\mathcal{C}$,
- the convex set $\mathcal{C} = V^\perp \cap \mathcal{D} \subset \mathbb{R}^M$ is compact as the intersection of the closed set $V^\perp$ and the compact set $\mathcal{D} \overset{\text{def.}}{=} \{\mathbf{c} \in \mathbb{R}^M : \left\|\boldsymbol{\nu}^*_{\mathcal{M}}(\mathbf{c})\right\|_\infty \leq 1\}$. The main argument to prove the compactness of $\mathcal{D}$ is that $\text{Im}(\boldsymbol{\nu}^*_{\mathcal{M}}) \subset \mathcal{C}_0(\mathbb{R})$ is finite dimensional. Let us prove it in a formal way. Consider the map $F : \mathbb{R}^M \to \mathcal{F}$ given by

$$\forall \mathbf{c} \in \mathbb{R}^M, \quad F(\mathbf{c}) \overset{\text{def.}}{=} \sum_{m=1}^{M} c_m g(x_m, \cdot) = \boldsymbol{\nu}^*_{\mathcal{M}}(\mathbf{c}) \tag{B.22}$$

  (using (B.4) for the last equality), where $\mathcal{F} \overset{\text{def.}}{=} \text{Span}\,(\{g(x_m, \cdot) : 1 \leq m \leq M\})$. Then, F is

  - linear;
  - injective and thus bijective due to the linear independence of the family $(g(x_m, \cdot))_{1\leq m\leq M}$. This independence can be proved by considering that $(g(x_m, \cdot))_{1\leq m\leq M}$ is a family of piecewise-linear splines with each finitely many knots, and so there exists a nonempty interval I in which all the $g(x_m, \cdot)$ are linear functions;
  - continuous with $\mathcal{F} \subset \mathcal{C}_0(\mathbb{R})$ endowed with the uniform norm $\|\cdot\|_\infty$.

Therefore, by the bounded inverse theorem, F^{-1} is continuous. Moreover, note that $\mathcal{E} \overset{\text{def.}}{=} \{f \in \mathcal{F} : \|f\|_\infty \leq 1\}$ is bounded and closed, and is thus compact (since $\mathcal{F} = \mathrm{Im}(\boldsymbol{\nu}^*_{\mathcal{M}})$ is finite dimensional). This proves that $\mathcal{D} = F^{-1}(\mathcal{E})$ is compact.

The convexity and the compactness of $\mathcal{C}$ imply that there is at least one extreme point of $\mathcal{C}$ that is a solution of problem (B.5). Hence, the assumptions of [97, Proposition 2.4, Chapter 3] are satisfied, which implies that any solution $(w, (\beta_0, \beta_1)) \in \mathcal{M}(\mathbb{R}) \times \mathbb{R}^2$ of (the primal) problem (B.2) and $\mathbf{c} \in \mathbb{R}^M$ of (the dual) problem (B.5) are linked by the optimality conditions

$$\boldsymbol{\nu}_{\mathcal{M}}(w) + \beta_0 \mathbf{1} + \beta_1 \mathbf{x} = \mathbf{y}_0, \tag{B.23}$$

$$\langle \mathbf{c},\ \mathbf{1} \rangle = \langle \mathbf{c},\ \mathbf{x} \rangle = 0, \quad \|w\|_{\mathcal{M}} = \left\langle w,\ \boldsymbol{\nu}^*_{\mathcal{M}}(\mathbf{c}) \right\rangle \quad \text{and} \quad \left\| \boldsymbol{\nu}^*_{\mathcal{M}}(\mathbf{c}) \right\|_\infty \leq 1. \tag{B.24}$$

Conversely, if any $(w, (\beta_0, \beta_1)) \in \mathcal{M}(\mathbb{R}) \times \mathbb{R}^2$ and $\mathbf{c} \in \mathbb{R}^M$ satisfy the optimality conditions given above, then again by [97, Proposition 2.4, Chapter 3] we obtain that $(w, (\beta_0, \beta_1)) \in \mathcal{M}(\mathbb{R}) \times \mathbb{R}^2$ and $\mathbf{c} \in \mathbb{R}^M$ are solutions of the primal and dual problems respectively. This proves the last statement of the lemma.

The last intermediate result needed for the proof of Proposition 1 is given in the next lemma, where we prove that any continuous function $\boldsymbol{\nu}^*_{\mathcal{M}}(\mathbf{c}) \in \mathcal{C}_0(\mathbb{R})$ with $\mathbf{c} \in \mathbb{R}^M$ satisfying the orthogonality conditions given in (B.7) is a piecewise-linear spline whose knots are located at the sampling points $\mathbf{x} = (x_m)_{1 \leq m \leq M}$.

Lemma 4. *Let $\mathbf{c} \in \mathbb{R}^M$ such that $\langle \mathbf{c},\ \mathbf{1} \rangle = \langle \mathbf{c},\ \mathbf{x} \rangle = 0$. Then, we have $\boldsymbol{\nu}^*_{\mathcal{M}}(\mathbf{c}) = \sum_{m=1}^M c_m (x_m - \cdot)_+$.*

Proof. We know by (A.5) and (B.4) that

$$\boldsymbol{\nu}^*_{\mathcal{M}}(\mathbf{c}) = \left\langle \mathbf{c},\ (g(x_m, \cdot))_{1 \leq m \leq M} \right\rangle, \tag{B.25}$$

$$= \left\langle \mathbf{c},\ ((x_m - x)_+ - (-x)_+ + x_m((-x)_+ - (1-x)_+))_{1 \leq m \leq M} \right\rangle, \tag{B.26}$$

$$= \left\langle \mathbf{c},\ ((x_m - x)_+)_{1 \leq m \leq M} \right\rangle - (-x)_+ \underbrace{\langle \mathbf{c},\ \mathbf{1} \rangle}_{=0} + ((-x)_+ - (1-x)_+) \underbrace{\langle \mathbf{c},\ \mathbf{x} \rangle}_{=0}, \tag{B.27}$$

which proves that $\boldsymbol{\nu}^*_{\mathcal{M}}(\mathbf{c}) = \sum_{m=1}^M c_m (x_m - \cdot)_+$.

We can now prove Proposition 1.

Proof of Proposition 1. Suppose that $f_{\text{opt}} \in \mathrm{BV}^{(2)}(\mathbb{R})$ is a solution of (g-BPC). Then, f_{opt} satisfies the interpolation conditions $f_{\text{opt}}(x_m) = y_{0,m}$ for all $m \in \{1, \dots, M\}$, and $(w, (\beta_0, \beta_1)) \in \mathcal{M}(\mathbb{R}) \times \mathbb{R}^2$ is a solution of problem (B.2) where $f_{\text{opt}} = \mathrm{D}_0^{-2}\{w\} + \beta_0 + \beta_1(\cdot)$. By Lemma 3, there exists a $\mathbf{c} \in \mathbb{R}^M$ solution of problem (B.5) which then satisfies $\langle \mathbf{c},\ \mathbf{1} \rangle = \langle \mathbf{c},\ \mathbf{x} \rangle = 0$ with $\left\| \boldsymbol{\nu}^*_{\mathcal{M}}(\mathbf{c}) \right\|_\infty \leq 1$. Let us denote $\eta \overset{\text{def.}}{=} \boldsymbol{\nu}^*_{\mathcal{M}}(\mathbf{c}) \in \mathcal{C}_0(\mathbb{R})$. By Lemma 4, we have $\eta = \sum_{m=1}^M c_m (x_m - \cdot)_+$ *i.e.*, η is a dual pre-certificate (Definition 2). Moreover, again by Lemma 3, we know that $\|w\|_{\mathcal{M}} = \langle w,\ \eta \rangle$ which gives the direct implication.

For the reverse implication, the dual pre-certificate η given by the statement satisfies $\eta = \boldsymbol{\nu}^*_{\mathcal{M}}(\mathbf{c})$ by Lemma 4, and since f_{opt} satisfies the interpolation conditions, we deduce that $\boldsymbol{\nu}_{\mathcal{M}}(w) + \beta_0 \mathbf{1} + \beta_1 \mathbf{x} = \mathbf{y}_0$ where β_0 and β_1 are defined thanks to the relation $f_{\text{opt}} = \mathrm{D}_0^{-2}\{w\} + \beta_0 + \beta_1(\cdot)$. Hence, by Lemma 3, $(w, (\beta_0, \beta_1)) \in \mathcal{M}(\mathbb{R}) \times \mathbb{R}^2$ is a solution of problem (B.2) (and $\mathbf{c}$ is a solution of problem (B.5)), *i.e.*, f_{opt} is a solution of (g-BPC).

Let us now prove that the relation $\|w\|_{\mathcal{M}} = \langle w,\ \eta \rangle$ is equivalent to $\mathrm{supp}_\pm(w) \subset \mathrm{sat}_\pm(\eta)$ when η is a dual pre-certificate (see Definition 3 for the definition of the signed support and signed saturation set). First, we have that $\|w\|_{\mathcal{M}} = \left\| w_{|\,\mathrm{sat}_+(\eta)} \right\|_{\mathcal{M}} + \left\| w_{|\,\mathrm{sat}_-(\eta)} \right\|_{\mathcal{M}} + \left\| w_{|S^c} \right\|_{\mathcal{M}}$ (see [98, Theorem 6.2]), where $S \overset{\text{def.}}{=} \mathrm{sat}_+(\eta) \cup \mathrm{sat}_-(\eta)$, hence

$$\left(\left\| w_{|\,\mathrm{sat}_+(\eta)} \right\|_{\mathcal{M}} - \left\langle w_{|\,\mathrm{sat}_+(\eta)},\ \eta \right\rangle \right) + \left(\left\| w_{|\,\mathrm{sat}_-(\eta)} \right\|_{\mathcal{M}} - \left\langle w_{|\,\mathrm{sat}_-(\eta)},\ \eta \right\rangle \right) + \left(\left\| w_{|S^c} \right\|_{\mathcal{M}} - \left\langle w_{|S^c},\ \eta \right\rangle \right) = 0. \tag{B.28}$$

Each of the three terms in the sum is nonnegative by definition of $\|\cdot\|_{\mathcal{M}}$, and the fact that $\|\eta\|_\infty \leq 1$, so that the equality $\|w\|_{\mathcal{M}} = \langle w,\ \eta \rangle$ is equivalent to

$$\left\| w_{|\,\mathrm{sat}_+(\eta)} \right\|_{\mathcal{M}} = \left\langle w_{|\,\mathrm{sat}_+(\eta)},\ \eta \right\rangle, \tag{B.29}$$

$$\left\| w_{|\,\mathrm{sat}_-(\eta)} \right\|_{\mathcal{M}} = \left\langle w_{|\,\mathrm{sat}_-(\eta)},\ \eta \right\rangle, \tag{B.30}$$

$$\left\| w_{|S^c} \right\|_{\mathcal{M}} = \left\langle w_{|S^c},\ \eta \right\rangle. \tag{B.31}$$

Consider the Jordan decomposition of w: $w = w_+ - w_-$. Then $\left\| w_{|\,\mathrm{sat}_+(\eta)} \right\|_{\mathcal{M}} = w_+(\mathrm{sat}_+(\eta)) + w_-(\mathrm{sat}_+(\eta))$ and $\left\langle w_{|\,\mathrm{sat}_+(\eta)},\ \eta \right\rangle = \int_{\mathrm{sat}_+(\eta)} \mathrm{d}w = w_+(\mathrm{sat}_+(\eta)) - w_-(\mathrm{sat}_+(\eta))$, so that (B.29) is equivalent to $w_-(\mathrm{sat}_+(\eta)) = 0$ *i.e.*,

$$\mathrm{supp}(w_-) \cap \mathrm{sat}_+(\eta) = \varnothing. \tag{B.32}$$

Similarly, we can prove that (B.30) is equivalent to

$$\mathrm{supp}(w_+) \cap \mathrm{sat}_-(\eta) = \varnothing, \tag{B.33}$$

since $\langle w_{|\operatorname{sat}_-(\eta)}, \eta\rangle = -\int_{\operatorname{sat}_-(\eta)} \mathrm{d}w$. As a result, to obtain the desired equivalence, it remains to prove that (B.31) is the same as $w_{|S^c} = 0$. The arguments can be found for example in [5] (see the proof of Lemma A.1), but we reproduce the reasoning here for the sake of completeness. Consider the closed sets for all $k > 0$

$$\Omega_k \overset{\text{def.}}{=} \mathbb{R} \setminus \left(S + \left(-\frac{1}{k}, \frac{1}{k} \right) \right) \subset S^c. \tag{B.34}$$

Suppose by contradiction that there exists $k > 0$ such that $\|w_{|\Omega_k}\|_{\mathcal{M}} > 0$. Since $|\eta| < 1$ on the closed set Ω_k (because it is true on the bigger open set S^c), we deduce that $\langle w_{|\Omega_k}, \eta\rangle < \|w_{|\Omega_k}\|_{\mathcal{M}}$ and then

$$\|w\|_{\mathcal{M}} = \langle w_{|\Omega_k}, \eta\rangle + \left\langle w_{|\Omega_k^c}, \eta \right\rangle < \|w_{|\Omega_k}\|_{\mathcal{M}} + \left\| w_{|\Omega_k^c} \right\|_{\mathcal{M}} = \|w\|_{\mathcal{M}}, \tag{B.35}$$

which is a contradiction. Hence, we have $\|w_{|\Omega_k}\|_{\mathcal{M}} = 0$ for all $k > 0$, which yields $\|w_{|S^c}\|_{\mathcal{M}} = 0$ since $S^c = \cup_{k>0} \Omega_k$, *i.e.*, $w_{|S^c} = 0$.

Appendix C. Proof of Proposition 2

The proof of Proposition 2 is very similar to the proof of Proposition 1, and is derived from the optimality conditions given in Lemma 3.

Proof of Proposition 2. Let η be a dual certificate in the sense of Proposition 2. By definition of η (it is in particular a dual pre-certificate in the sense of Definition 2) and by Lemma 4, there exists $\mathbf{c} \in \mathbb{R}^M$ such that $\eta = \boldsymbol{\nu}_{\mathcal{M}}^*(\mathbf{c})$ and $\langle \mathbf{c}, \mathbf{1}\rangle = \langle \mathbf{c}, \mathbf{x}\rangle = 0$. Since η is a dual certificate, Proposition 1 implies that there exists a $\tilde{f} \in \mathrm{BV}^{(2)}(\mathbb{R})$ satisfying the interpolation conditions and such that $\left\|\mathrm{D}^2\tilde{f}\right\|_{\mathcal{M}} = \left\langle \mathrm{D}^2\tilde{f}, \eta \right\rangle$. This implies that $\mathbf{c}$ and $(\tilde{w}, (\tilde{\beta}_0, \tilde{\beta}_1)) \in \mathcal{M}(\mathbb{R}) \times \mathbb{R}^2$, where $\tilde{f} = \mathrm{D}_0^{-2}\{\tilde{w}\} + \tilde{\beta}_0 + \tilde{\beta}_1(\cdot)$, satisfy (B.6) and (B.7) *i.e.*, in particular $\mathbf{c}$ is a solution of the dual problem (B.5) by Lemma 3. Using this fixed vector $\mathbf{c} \in \mathbb{R}^M$ and the decomposition of any $f \in \mathrm{BV}^{(2)}(\mathbb{R})$ as $f = \mathrm{D}_0^{-2}\{w\} + \beta_0 + \beta_1(\cdot)$ (see (7)), the equivalence in Lemma 3 directly yields that f_{opt} is a solution of (g-BPC) if and only if f_{opt} satisfies the interpolation conditions $f_{\text{opt}}(x_m) = y_{0,m}$ and $\|\mathrm{D}^2 f_{\text{opt}}\|_{\mathcal{M}} = \langle \mathrm{D}^2 f_{\text{opt}}, \eta\rangle$, which concludes the proof.

Appendix D. Proof of Proposition 3

Let $f_{\text{opt}} \in \mathrm{BV}^{(2)}(\mathbb{R})$ be a solution of problem (g-BPC) given by Theorem 1. By (7), there exist $w \in \mathcal{M}(\mathbb{R})$ and $(\beta_0, \beta_1) \in \mathbb{R}^2$ such that $f_{\text{opt}} = \mathrm{D}_0^{-2}\{w\} + \beta_0 + \beta_1(\cdot)$. By the assumption of the proposition, there exists a nondegenerate dual certificate η, so that by applying Proposition 2, we obtain $\operatorname{supp}_{\pm}(w) \subset \operatorname{sat}_{\pm}(\eta)$. Moreover, we have that $\operatorname{sat}_{\pm}(\eta) \subset \{x_2, \ldots, x_{M-1}\}$ due to the two following facts

- $\eta = \sum_{m=1}^{M} c_m (x_m - \cdot)_+$ (as a dual pre-certificate, see Lemma 4),
- $\operatorname{sat}_{\pm}(\eta)$ is a discrete set (as η is nondegenerate).

This implies that η must be equal to ± 1 at the points $\{x_2, \ldots, x_{M-1}\}$, which yields

$$w = \sum_{k=2}^{M-1} a_k \delta(\cdot - x_k), \tag{D.1}$$

where the $a_k \in \mathbb{R}$ are (possibly zero) weights. In particular, this implies that f_{opt} is a piecewise-linear spline with at most $(M-2)$ knots that are a subset of $\{x_2, \ldots, x_{M-1}\}$. It remains to prove that the coefficients $a_2, \ldots, a_{M-1}, \beta_0, \beta_1$ are uniquely determined to conclude that f_{opt} is the unique solution of (g-BPC).

Since f_{opt} is a solution of (g-BPC), we have that $\boldsymbol{\nu}(f_{\text{opt}}) = \mathbf{y}_0$. This implies that

$$\sum_{k=2}^{M-1} a_k \mathbf{g}_k + \beta_0 \mathbf{1} + \beta_1 \mathbf{x} = \mathbf{y}_0 \quad \text{with} \quad \mathbf{g}_k \overset{\text{def.}}{=} \boldsymbol{\nu}_{\mathcal{M}}(\delta(\cdot - x_k)) = (g(x_m, x_k))_{1 \le m \le M} \in \mathbb{R}^M. \tag{D.2}$$

We now prove that this equation uniquely determines the coefficients $a_2, \ldots, a_{M-1}, \beta_0, \beta_1$ by showing that the family $(\mathbf{1}, \mathbf{x}, \mathbf{g}_2, \ldots, \mathbf{g}_{M-1})$ is a basis of $\mathbb{R}^M$. Indeed, by definition of g (see (A.5)), we have that

$$\forall k \in \{2, \ldots, M-1\}, \quad \mathbf{g}_k = ((x_m - x_k)_+)_{1 \le m \le M} - (-x_k)_+ \mathbf{1} + ((-x_k)_+ - (1 - x_k)_+)\, \mathbf{x}. \tag{D.3}$$

Hence, by writing the matrix of the family $(\mathbf{1}, \mathbf{x}, \mathbf{g}_2, \ldots, \mathbf{g}_{M-1})$ in the canonical basis of $\mathbb{R}^M$, subtracting thanks to (D.3) appropriate linear combinations of the first two columns (given by the vectors $\mathbf{1}$ and $\mathbf{x}$) to all of the other columns and

finally subtracting x_1 times the first column to the second one, we end up with the following matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & (x_2 - x_1) & 0 & 0 & \dots & 0 \\ 1 & (x_3 - x_1) & (x_3 - x_2) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (x_M - x_1) & (x_M - x_2) & (x_M - x_3) & \dots & (x_M - x_{M-1}) \end{pmatrix}. \tag{D.4}$$

The latter is a lower triangular matrix with nonzero coefficients on the diagonal (as the sampling points x_m are pairwise distinct), and is thus invertible, which proves the desired result.

Appendix E. Proof of Theorem 3

Let $f_{\text{opt}} \in \mathcal{V}_0$. We fix $m \in \{2, M-2\}$. First of all, as we have seen in the proof of Theorem 2, if $a_m a_{m+1} \leq 0$, then $f_{\text{opt}} = f_{\text{cano}}$ on $[x_m, x_{m+1}]$, and the graph of f_{opt} in this interval is equal to the one of f_{cano}. Assume now that $a_m a_{m+1} > 0$. We now show that $\{(x, f_{\text{opt}}(x)) : x \in [x_m, x_{m+1}]\} \subset \Delta_m$. The slope condition $a_m a_{m+1} > 0$ implies that η_{cano} is degenerate and that $\eta_{\text{cano}} = \pm 1$ is constant over $[x_m, x_{m+1}]$. Assume for instance that the value is 1, in which case f_{opt} is convex over $[x_{m-1}, x_{m+2}]$ according to Theorem 2.

We shall use the following well-known fact on convex functions. Fix $a < b < c$ and assume that f is convex over $[a, c]$. Then, f is below its arc between a and b on (a, b), that is, $f(x) \leq \frac{f(b)-f(a)}{b-a}(x-a) + f(a)$ for any $x \in (a, b)$. Moreover, f is above the same arc over (b, c), that is, $f(x) \geq \frac{f(b)-f(a)}{b-a}(x-a) + f(a)$ for any $x \in (b, c)$.

Let $x^* \in [x_m, x_{m+1}]$. By convexity, f_{opt} is below its arc between x_m and x_{m+1}. Hence we have that

$$f_{\text{opt}}(x^*) \leq \frac{y_{0,m+1} - y_{0,m}}{x_{m+1} - x_m}(x^* - x_m) + y_{0,m}. \tag{E.1}$$

Moreover, the convexity over $[x_{m-1}, x^*]$ implies that $f_{\text{opt}}(x^*)$ is above the arc of f_{opt} between x_{m-1} and x_m. This implies that

$$f_{\text{opt}}(x^*) \geq \frac{y_{0,m} - y_{0,m-1}}{x_m - x_{m-1}}(x^* - x_{m-1}) + y_{0,m-1}. \tag{E.2}$$

A similar argument over $[x^*, x_{m+2}]$ implies that

$$f_{\text{opt}}(x^*) \geq \frac{y_{0,m+2} - y_{0,m+1}}{x_{m+2} - x_{m+1}}(x^* - x_{m+1}) + y_{0,m+1}. \tag{E.3}$$

The conditions (E.1), (E.2), and (E.3) are precisely equivalent to $(x^*, f_{\text{opt}}(x^*)) \in \Delta_m$, since the three linear equations delineate this domain in this case. The same proof applies when $\eta_{\text{cano}} = -1$ over $[x_m, x_{m+1}]$ by using concavity instead of convexity. This proves that $\mathcal{G}(f_{\text{opt}}) \subset \mathcal{G}(f_{\text{cano}}) \cup (\cup_{m \in \mathcal{X}} \Delta_m)$ for every $f_{\text{opt}} \in \mathcal{V}_0$, and hence the direct inclusion in (30).

For the reverse inclusion, we already know that $f_{\text{cano}} \in \mathcal{V}_0$, therefore it suffices to show that, for any $m \in \mathcal{X}$ and any $(x^*, y^*) \in \Delta_m$, there exists a solution $f_{\text{opt}} \in \mathcal{V}_0$ such that $f_{\text{opt}}(x^*) = y^*$. As before, since $m \in \mathcal{X}$, we know that $\eta_{\text{cano}} = \pm 1$ on $[x_m, x_{m+1}]$ and we can assume without loss of generality that the value is 1. Then, any solution is convex and satisfies the relations (E.1), (E.2), and (E.3). By convexity of $\mathcal{V}_0$, it suffices to show the result for (x^*, y^*) in the boundary of Δ_m, which is delimited by the relations

$$\frac{y_{0,m+1} - y_{0,m}}{x_{m+1} - x_m}(x^* - x_m) + y_{0,m} = y^*, \text{ or} \tag{E.4}$$

$$\frac{y_{0,m} - y_{0,m-1}}{x_m - x_{m-1}}(x^* - x_{m-1}) + y_{0,m-1} = y^*, \text{ or} \tag{E.5}$$

$$\frac{y_{0,m+2} - y_{0,m+1}}{x_{m+2} - x_{m+1}}(x^* - x_{m+1}) + y_{0,m+1} = y^*. \tag{E.6}$$

The solution f_{cano} is such that $f_{\text{cano}}(x^*) = \frac{y_{0,m+1} - y_{0,m}}{x_{m+1} - x_m}(x^* - x_m) + y_{0,m} = y^*$, hence any (x^*, y^*) satisfying (E.4) is attained by a solution (the canonical one) in $\mathcal{V}_0$. Assume that (x^*, y^*) satisfies (E.5) (the case of (E.6) follows the same argument). We construct f_{opt} as follows. First, $f_{\text{opt}}(x) = f_{\text{cano}}(x)$ for any $x \notin (x_m, x_{m+1})$. Then, we set

$$f_{\text{opt}}(x) = \frac{y_{0,m} - y_{0,m-1}}{x_m - x_{m-1}}(x - x_{m-1}) + y_{0,m-1} \tag{E.7}$$

for $x \in (x_m, x^*]$. In particular, $f(x^*) = y^*$, and f_{opt} is linear on $[x_m, x^*]$. Finally, we impose that f_{opt} is linear on $[x^*, x_{m+1}]$, which is equivalent to the relation

$$f_{\text{opt}}(x) = \frac{y_{0,m+1} - y^*}{x_{m+1} - x^*}(x - x^*) + y^* \tag{E.8}$$

for any $x \in [x^*, x_{m+1}]$. We then claim that $f_{\text{opt}} \in \mathcal{V}_0$, the argument being very similar to the one of Lemma 1. Indeed, to show this, it suffices to remark that f_{opt}, which is piecewise-constant and coincides with f_{cano} outside of (x_m, x_{m+1}), is

convex on $[x_{m-1}, x_{m+2}]$ (this is guaranteed by the slope condition $a_m a_{m+1} > 0$ and the construction of f_{opt}). According to Theorem 2, this implies that $f_{\text{opt}} \in \mathcal{V}_0$, with $f_{\text{opt}}(x^*) = y^*$. This finally shows that $(x^*, y^*) \in \cup_{f_{\text{opt}} \in \mathcal{V}_0} \mathcal{G}(f_{\text{opt}})$, which proves (30).

Appendix F. Proof of Theorem 4

Using Theorem 2, for any $f_{\text{opt}} \in \mathcal{V}_0$, we have $f_{\text{opt}}(x) = f_{\text{cano}}(x)$ for any x such that $\eta_{\text{cano}}(x) \neq \pm 1$. We now focus on regions where $\eta_{\text{cano}}(x) = \pm 1$. For all $n \in \{1, \dots, N_s\}$, f_{cano} has $\alpha_n + 1$ knots in the interval $[x_{s_n}, x_{s_n+\alpha_n}]$. In order to construct one of the sparsest solutions, we must therefore replace these $\alpha_n + 1$ knots with as little knots as possible in each saturation region, since all solutions must coincide with f_{cano} outside these regions. In order to lighten the notations, in what follows, we focus on a single saturation region determined by a fixed $n \in \{1, \dots, N_s\}$ and we write $\alpha \overset{\text{def.}}{=} \alpha_n$ and $s \overset{\text{def.}}{=} s_n$.

Similarly to the proof of Proposition 6, a piecewise-linear spline f that coincides with f_{cano} outside the interval $[x_s, x_{s+\alpha}]$ must be of the form

$$f(x) = f_{\text{cano}}(x) - \sum_{n'=0}^{\alpha} a_{s+n'}(x - x_{s+n'})_+ + \sum_{p=1}^{P} \tilde{a}_p (x - \tilde{\tau}_p)_+, \tag{F.1}$$

where $\tilde{a}_p \in \mathbb{R}$, $\tilde{\tau}_p \in [x_s, x_{s+\alpha}]$ such that $\tilde{\tau}_1 < \dots < \tilde{\tau}_P$ and P is the number of knots of f in this interval. We then prove the following lemma.

Lemma 5. *If f in (F.1) satisfies the constraints $f(x_m) = y_{0,m}$ for all $m \in \{1, \dots, M\}$, then the number of knots P in $[x_s, x_{s+\alpha}]$ satisfies $P \geq \lceil \frac{\alpha+1}{2} \rceil$.*

Proof. Lemma 5 is trivially true for $\alpha = 0$, since we must have $f = f_{\text{cano}}$ and thus $P = 1$. Assume now that $\alpha > 0$. Firstly, we show that we must have $\tilde{\tau}_1 \in [x_s, x_{s+1})$. Assume by contradiction that $\tilde{\tau}_1 \geq x_{s+1}$: then, f has no knots in the interval (x_{s-1}, x_{s+1}). Yet f must satisfy the interpolation constraints $f(x_m) = y_{0,m}$ for all $m \in \{1, \dots, M\}$, which implies that the points $\mathrm{P}_{0,s-1}$, $\mathrm{P}_{0,s}$, and $\mathrm{P}_{0,s+1}$ are aligned. Therefore, f_{cano} has a weight $a_s = 0$ (defined in (21)) which implies that $\eta_{\text{cano}}(x_s) = 0$, which contradicts the assumption $\eta_{\text{cano}}(x_s) = \pm 1$. We can then prove in a similar fashion that $\tilde{\tau}_P \in (x_{s+\alpha-1}, x_{s+\alpha}]$ when $\alpha > 1$.

Next, we show that for $\alpha \geq 2$, we have

$$\forall n' \in \{1, \dots, \alpha - 1\},\ \exists p \in \{1, \dots, P\} \text{ such that } \tilde{\tau}_p \in (x_{s+n'-1}, x_{s+n'+1}), \tag{F.2}$$

i.e., there must be a knot in all blocks of two consecutive saturation intervals. We assume by contradiction that this is not the case. Similarly to above, this implies that $\mathrm{P}_{0,s+n'-1}$, $\mathrm{P}_{0,s+n'}$, and $\mathrm{P}_{0,s+n'+1}$ are aligned and thus that $\eta_{\text{cano}}(x_{s+n'}) = 0$, which yields a contradiction.

Lemma 5 immediately follows from the constraints $\tilde{\tau}_1 \in [x_s, x_{s+1})$ and $\tilde{\tau}_P \in [x_{s+\alpha-1}, x_{s+\alpha}]$ for $\alpha \leq 2$. For $\alpha > 2$, by the two aforementioned constraints, f must have at least two knots in the first and last saturation intervals $[x_s, x_{s+1})$ and $(x_{s+\alpha-1} x_{s+\alpha}]$ respectively. Next, consider the interval $[x_{s+1}, x_{s+\alpha-1}]$, which consists of the central $\alpha - 2$ consecutive saturations. Using (F.2), this interval must contain at least $\lfloor \frac{\alpha-2}{2} \rfloor$ knots, which yields the lower bound $P \geq 2 + \lfloor \frac{\alpha-2}{2} \rfloor = \lceil \frac{\alpha+1}{2} \rceil$ (the last equality can easily be verified for every $\alpha \in \mathbb{N}$).

The following Lemma then states that the bound in Lemma 5 is tight.

Lemma 6. *The lower bound in Lemma 5 is always reached,* i.e., *there exists a piecewise-linear spline $f_{\text{opt}} \in \mathcal{V}_0$ of the form (F.1) with $P = \lceil \frac{\alpha+1}{2} \rceil$ knots in $[x_s, x_{s+\alpha}]$. If α is odd or $\alpha = 0$, then f_{opt} is unique. If $\alpha > 0$ is even, then there are uncountably many such functions f_{opt}.*

Proof. Lemma 6 is trivially true for $\alpha = 0$, *i.e.*, when no saturation occurs. Indeed, the saturation interval is then reduced to the point $\{x_s\}$, and the only solution $f_{\text{opt}} \in \mathcal{V}_0$ of the form (F.1) is $f_{\text{opt}} = f_{\text{cano}}$ for which $P = 1$.

Assume now that $\alpha = 2k+1$ is odd. The bound in Lemma 5 then reads $P \geq k+1$. Similarly to the proof of Proposition 6, we construct a function f_{opt} of the form (F.1) with $P = k + 1$ and

$$\begin{cases} \tilde{a}_1 \overset{\text{def.}}{=} a_s + a_{s+1} \text{ and } \tilde{\tau}_1 \overset{\text{def.}}{=} \frac{a_s x_s + a_{s+1} x_{s+1}}{\tilde{a}_1}; \\ \tilde{a}_2 \overset{\text{def.}}{=} a_{s+2} + a_{s+3} \text{ and } \tilde{\tau}_2 \overset{\text{def.}}{=} \frac{a_{s+2} x_{s+2} + a_{s+3} x_{s+3}}{\tilde{a}_2}; \\ \vdots \\ \tilde{a}_{k+1} \overset{\text{def.}}{=} a_{s+2k} + a_{s+2k+1} \text{ and } \tilde{\tau}_k \overset{\text{def.}}{=} \frac{a_{s+2k} x_{s+2k} + a_{s+2k+1} x_{s+2k+1}}{\tilde{a}_{k+1}}. \end{cases} \tag{F.3}$$

Since the $a_s, \dots, a_{s+\alpha}$ all have the same (nonzero) sign, the $\tilde{\tau}_i$, $i = 1, \dots, k+1$, are all barycenters with positive weights, which implies that $\tilde{\tau}_i \in (x_{s+2i}, x_{s+2i+1})$. Then, as in the proof of Proposition 6, replacing the knots at x_{s+2i} and x_{s+2i+1} in

f_{cano} by a single knot at $\tilde{\tau}_i$ does not change the expression of f_{opt} outside the interval (x_{s+2i}, x_{s+2i+1}), which implies that all the constraints $f_{\text{opt}}(x_m) = y_{0,m}$ for all $m \in \{1, \dots, M\}$ are satisfied.

Next, let $I_s = \{1, \dots, M\} \setminus \{s, \dots, s+\alpha\}$ be the set of indices outside our interval of interest. Since $a_s, \dots, a_{s+\alpha}$ and thus $\tilde{a}_1, \dots, \tilde{a}_{k+1}$ all have the same sign, we have $\|\mathrm{D}^2 f_{\text{opt}}\|_{\mathcal{M}} = \sum_{m\in I_s} |a_m| + |\sum_{i=1}^{k+1} \tilde{a}_i| = \sum_{m\in I_s} |a_m| + |\sum_{n=0}^{\alpha} a_{s+n}| = \|\mathrm{D}^2 f_{\text{cano}}\|_{\mathcal{M}}$, which together with the interpolation constraints implies that $f_{\text{opt}} \in \mathcal{V}_0$.

To show the uniqueness, consider once again a function f_{opt} of the form (F.1) with $P = k+1$ and $\tilde{\tau}_1 < \cdots < \tilde{\tau}_{k+1}$. We then invoke Lemma 5, which stipulates that there must be knots in the first and last saturation intervals as well as every two consecutive saturation intervals. The only way to achieve this is to have $\tilde{\tau}_i \in (x_{s+2i}, x_{s+2i+1})$, $i = 0, \dots, k$. The intervals (x_{s+2i-1}, x_{s+2i}) for all $i \in \{1, \dots, k\}$ thus have no knots, which implies that in these intervals, f_{opt} must follow the line $(\mathrm{P}_{0,s+2i-1}, \mathrm{P}_{0,s+2i})$. The knots are then necessarily the intersection of these lines, which yields the solution given in (F.3). The latter is therefore the unique function in $\mathcal{V}_0$ with $P = k+1$ knots in the interval $[x_s, x_{s+\alpha}]$. An example of such a sparsest solution is shown in Fig. 5 with $M = 6$ and $\alpha = 3$ consecutive saturation intervals.

Assume now that $\alpha = 2k$ is even, with $k > 0$. The bound in Lemma 5 then reads $P \geq k+1$. By Lemma 1, the intersection $\widetilde{\mathrm{P}} = \begin{bmatrix}\tilde{\tau} & \tilde{y}\end{bmatrix}^T$ between the lines $(\mathrm{P}_{0,s-1}, \mathrm{P}_{0,s})$ and $(\mathrm{P}_{0,s+1}, \mathrm{P}_{0,s+2})$ exists and satisfies $\tilde{\tau} \in (x_s, x_{s+1})$. Then, let $\widetilde{\mathrm{P}}_1 = \begin{bmatrix}\tilde{\tau}_1 & \tilde{y}_1\end{bmatrix}^T$ be any point on the line segment $[\mathrm{P}_{0,s}, \widetilde{\mathrm{P}}]$, *i.e.*, with $\tilde{\tau}_1 \in [x_s, \tilde{\tau}]$. Then, we define $\widetilde{\mathrm{P}}_2$ as the intersection between the lines $(\widetilde{\mathrm{P}}_1, \mathrm{P}_{0,s+1})$ and $(\mathrm{P}_{0,s+2}, \mathrm{P}_{0,s+3})$. Similarly, if $\alpha \geq 4$, for every $i \in \{3, \dots, k+1\}$, we define $\widetilde{\mathrm{P}}_i = \begin{bmatrix}\tilde{\tau}_i & \tilde{y}_i\end{bmatrix}^T$ as the intersection between the lines $(\mathrm{P}_{0,s+2i-4}, \mathrm{P}_{0,s+2i-3})$ and $(\mathrm{P}_{0,s+2i-2}, \mathrm{P}_{0,s+2i-1})$. Due to a similar barycenter argument as in (F.3), these intersections are well defined and satisfy $\tilde{\tau}_i \in (x_{s+2i-3}, x_{s+2i-2})$. Let f_{opt} be the piecewise-linear spline that coincides with f_{cano} outside the interval $(x_s, x_{s+\alpha})$, and that connects the points $\mathrm{P}_{0,s-1}$, $\widetilde{\mathrm{P}}_1, \dots, \widetilde{\mathrm{P}}_{k+1}$, and $\mathrm{P}_{0,s+\alpha}$ in that interval. By construction, f_{opt} satisfies the constraints $f_{\text{opt}}(x_m) = y_{0,m}$, $m \in \{1, \dots, M\}$. Moreover, once again in a similar manner to (F.3), we have that $\|f_{\text{opt}}\|_{\mathcal{M}} = \|f_{\text{cano}}\|_{\mathcal{M}}$, which implies that $f_{\text{opt}} \in \mathcal{V}_0$. Finally, f_{opt} is of the form (F.1) with the lowest possible sparsity $P = k+1$ in the interval $[x_s, x_{s+\alpha}]$ (by Lemma 5). Yet there are uncountably many possible choices of $\widetilde{\mathrm{P}}_1$ (it can be any point on a non-singleton line segment). All of these choices lead to a different solution $f_{\text{opt}} \in \mathcal{V}_0$ that is uniquely defined, since the choice of $\widetilde{\mathrm{P}}_1$ specifies $\widetilde{\mathrm{P}}_2, \dots, \widetilde{\mathrm{P}}_{k+1}$. This proves that there are uncountably many solutions of the (g-BPC) with sparsity $k+1$ in $[x_s, x_{s+\alpha}]$, and that there is a single degree of freedom for the choice of these $k+1$ knots. An example of such a sparsest solution is shown in Fig. 6 with $M = 5$ and $\alpha = 2$ consecutive saturation intervals. In our algorithm, we simply choose $\widetilde{\mathrm{P}}_1 = \mathrm{P}_{0,s}$, which yields a function f_{opt} of the form (F.1) with

$$\begin{cases} \tilde{a}_1 \overset{\text{def.}}{=} a_s \text{ and } \tilde{\tau}_1 \overset{\text{def.}}{=} x_s; \\ \tilde{a}_2 \overset{\text{def.}}{=} a_{s+1} + a_{s+2} \text{ and } \tilde{\tau}_2 \overset{\text{def.}}{=} \frac{a_{s+1}x_{s+1} + a_{s+2}x_{s+2}}{\tilde{a}_2}; \\ \vdots \\ \tilde{a}_{k+1} \overset{\text{def.}}{=} a_{s+2k-1} + a_{s+2k} \text{ and } \tilde{\tau}_k \overset{\text{def.}}{=} \frac{a_{s+2k-1}x_{s+2k-1} + a_{s+2k}x_{s+2k}}{\tilde{a}_{k+1}}. \end{cases} \tag{F.4}$$

Theorem 4 then directly derives from Lemma 6 applied independently to each saturation interval $[x_{s_n}, x_{s_n+\alpha_n}]$ for $n \in \{1, \dots, N_s\}$. Note that Lemma 6 also applies when no saturation occurs, *i.e.*, $\alpha_n = 0$. A sparsest solution of the (g-BPC) thus coincides with a function of the form (F.1) constructed in Lemma 6 in each of these intervals, and with f_{cano} outside these intervals. Finally, since the behavior of a solution in each saturation interval does not affect its behavior outside of it, the number of degrees of freedom in the set of sparsest solutions of the (g-BPC) is simply the sum of the number of degrees of freedom in each saturation interval. Yet by Lemma 6, there are no degrees of freedom in intervals such that α_n is odd (a sparsest solution is uniquely determined on that interval), and there is one when α_n is even. Therefore, the total number of degrees of freedom of the set of sparsest solutions of the (g-BPC) is equal to the number of even values of α_n for $n \in \{1, \dots, N_s\}$.

Appendix G. Proof of Proposition 7

Assume by contradiction that there exist $f_1, f_2 \in \mathcal{V}_\lambda$ and $m_0 \in \{1, \dots, M\}$ such that $f_1(x_{m_0}) \neq f_2(x_{m_0})$, and let $f_\gamma = \gamma f_1 + (1-\gamma) f_2$, where $0 < \gamma < 1$. We then have

$$\begin{aligned} &\sum_{m=1}^{M} E(f_\gamma(x_m), y_m) + \lambda \|\mathrm{D}^2 f_\gamma\|_{\mathcal{M}} \\ &< \gamma \sum_{m=1}^{M} E(f_1(x_m), y_m) + (1-\gamma) \sum_{m=1}^{M} E(f_2(x_m), y_m) + \lambda\Big(\gamma \|\mathrm{D}^2 f_1\|_{\mathcal{M}} + (1-\gamma)\|\mathrm{D}^2 f_2\|_{\mathcal{M}}\Big) \\ &= \gamma \mathcal{J}_\lambda + (1-\gamma)\mathcal{J}_\lambda = \mathcal{J}_\lambda, \end{aligned} \tag{G.1}$$

where $\mathcal{J}_\lambda$ is the optimal cost of the (g-BLASSO). The inequality is due to the convexity of the $\|\cdot\|_{\mathcal{M}}$ norm and of $E(\cdot, y)$ for any $y \in \mathbb{R}$. The fact that it is strict is due to the strict convexity of $E(\cdot, y_{m_0})$ and the fact that $f_1(x_{m_0}) \neq f_2(x_{m_0})$. Yet since $\mathcal{V}_\lambda$ is a convex set, we have $f_\gamma \in \mathcal{V}_\lambda$: this implies that $\mathcal{J}_\lambda = \sum_{m=1}^{M} E(f_\gamma(x_m), y_m) + \lambda \|\mathrm{D}^2 f_\gamma\|_{\mathcal{M}} < \mathcal{J}_\lambda$, which yields a contradiction.

Therefore, there exists a unique vector $\mathbf{y}_\lambda \in \mathbb{R}^M$ such that for any $f_{\text{opt}} \in \mathcal{V}_\lambda$, $f_{\text{opt}}(x_m) = y_{\lambda,m}$ for all $m \in \{1, \dots, M\}$. This implies that $\mathcal{V}_\lambda \subset \{f \in \text{BV}^{(2)}(\mathbb{R}) : f(x_m) = y_{\lambda,m},\ 1 \leq m \leq M\}$. Moreover, we have that for any $f_{\text{opt}} \in \mathcal{V}_\lambda$, $E(f_{\text{opt}}(x_m), y_m) = E(y_{\lambda,m}, y_m)$, and thus that the data fidelity is constant in the constrained space $\{f \in \text{BV}^{(2)}(\mathbb{R}) : f(x_m) = y_{\lambda,m},\ 1 \leq m \leq M\}$. This proves the equality between the solution sets of the (g-BLASSO) and (35).

Appendix H. Proof of Proposition 10

Item 1. Let $J(\beta_0, \beta_1) = \sum_{m=1}^M E(\beta_0 + \beta_1 x_m, y_m)$ be the objective function of problem (39). We show that problem (39) indeed has a unique solution by proving that J is strictly convex and coercive when $M \geq 2$ and the x_m are pairwise distinct.

Concerning the coercivity, let $\|(\beta_0, \beta_1)\|_2 \to +\infty$. Assume by contradiction that $\beta_0 + \beta_1 x_m$ is bounded for every $m \in \{1, \dots, M\}$. Then, since $M \geq 2$, $\beta_0 + \beta_1 x_1 - (\beta_0 + \beta_1 x_2) = \beta_1(x_1 - x_2)$ must also be bounded, which implies that β_1 is bounded since the x_m are pairwise distinct. Therefore, we must have $|\beta_0| \to +\infty$, which implies that $|\beta_0 + \beta_1 x_1| \to +\infty$ which yields a contradiction. Therefore, there exists a $m_0 \in \{1, \dots, M\}$ such that $|\beta_0 + \beta_1 x_{m_0}| \to +\infty$. The coercivity of J then directly follows from that of $E(\cdot, y_{m_0})$.

Next, to prove the strict convexity of J, let $(\beta_0, \beta_1), (\beta_0', \beta_1') \in \mathbb{R}^2$ with $(\beta_0, \beta_1) \neq (\beta_0', \beta_1')$, and $0 < s < 1$. For any m, we have $s\beta_0 + (1-s)\beta_0' + (s\beta_1 + (1-s)\beta_1')x_m = s(\beta_0 + \beta_1 x_m) + (1-s)(\beta_0' + \beta_1' x_m)$. Since $(\beta_0, \beta_1) \neq (\beta_0', \beta_1')$ and the x_m are distinct, the equation $\beta_0 + \beta_1 x_m = \beta_0' + \beta_1' x_m$ can only be satisfied for at most a single $m \in \{1, \dots, M\}$. Yet $M \geq 2$, which implies that $\exists m_0,\ \beta_0 + \beta_1 x_{m_0} \neq \beta_0' + \beta_1' x_{m_0}$. Therefore, due to the strict convexity of $E(\cdot, y_{m_0})$, we have

$$E((s\beta_0 + (1-s)\beta_0') + (s\beta_1 + (1-s)\beta_1')x_{m_0}, y_{m_0}) < sE(\beta_0 + \beta_1 x_{m_0}, y_{m_0}) + (1-s)E(\beta_0' + \beta_1' x_{m_0}, y_{m_0}). \tag{H.1}$$

It then follows from the convexity of $E(\cdot, y_m)$ for all m that $J\big(s(\beta_0, \beta_1) + (1-s)(\beta_0', \beta_1')\big) < sJ(\beta_0, \beta_1) + (1-s)J(\beta_0', \beta_1')$, which proves the strict convexity of J. Together with the fact that J is coercive, this proves that (39) has a unique solution.

Item 2. Assume that $\lambda \geq \lambda_{\max}$. By Fermat's rule, a vector $\mathbf{z}_{\text{opt}}$ is a solution of problem (36) if and only if the zero vector belongs to the subdifferential of the objective function evaluated at $\mathbf{z}_{\text{opt}}$. We thus have $\mathbf{z}_{\text{opt}} = \mathbf{y}_\lambda$ if and only if

$$\mathbf{0} \in \underbrace{\begin{pmatrix} \partial_1 E(z_{\text{opt},1}, y_1) \\ \vdots \\ \partial_1 E(z_{\text{opt},M}, y_M) \end{pmatrix}}_{\stackrel{\text{def.}}{=} \mathbf{v}(\mathbf{z}_{\text{opt}})} + \lambda \partial \|\mathbf{L} \cdot\|_1(\mathbf{z}_{\text{opt}}), \tag{H.2}$$

where ∂_1 denotes the partial derivative with respect to the first variable, and ∂ the subdifferential. The chain rule for subdifferentials [99, Theorem 23.9.] yields $\partial\|\mathbf{L}\cdot\|_1(\mathbf{z}) = \{\mathbf{L}^T\mathbf{g} : \mathbf{g} \in \partial\|\cdot\|_1(\mathbf{L}\mathbf{z}) \subset \mathbb{R}^{M-2}\}$, where $\partial\|\cdot\|_1(\mathbf{a}) = \{\mathbf{g} \in \mathbb{R}^{M-2} : \|\mathbf{g}\|_\infty \leq 1,\ \mathbf{a}^T\mathbf{g} = \|\mathbf{a}\|_1\}$. The vector $\mathbf{L}\mathbf{z}_{\text{opt}}$ lists the weights a_m associated to the knots of the canonical solution $f_{\mathbf{z}_{\text{opt}}}$ (see the proof of Proposition 8). Therefore, the linear regression case (in which $f_{\mathbf{z}_{\text{opt}}}$ has no knot) corresponds to $\mathbf{L}\mathbf{z}_{\text{opt}} = \mathbf{0}$. In this case, since $\partial\|\cdot\|_1(\mathbf{0}) = \{\mathbf{g} \in \mathbb{R}^{M-2} : \|\mathbf{g}\|_\infty \leq 1\}$, the optimality condition (H.2) now reads

$$\exists \mathbf{g} \in \mathbb{R}^{M-2}, \quad \|\mathbf{g}\|_\infty \leq 1, \quad \text{s.t.} \quad \mathbf{v}(\mathbf{z}_{\text{opt}}) + \lambda \mathbf{L}^T\mathbf{g} = \mathbf{0}. \tag{H.3}$$

We now prove that $\mathbf{z}_{\text{opt}} = \beta_{\text{opt}_0}\mathbf{1} + \beta_{\text{opt}_1}\mathbf{x}$ satisfies the optimality conditions (H.2), and thus that $\mathbf{y}_\lambda = \beta_{\text{opt}_0}\mathbf{1} + \beta_{\text{opt}_1}\mathbf{x}$. To achieve this, we prove that $\mathbf{g} = -\frac{1}{\lambda}\mathbf{L}^{T\dagger}\mathbf{v}(\mathbf{z}_{\text{opt}})$ satisfies $\mathbf{v}(\mathbf{z}_{\text{opt}}) + \lambda\mathbf{L}^T\mathbf{g} = \mathbf{0}$. Firstly, since $\lambda \geq \lambda_{\max}$, we have that $\|\mathbf{g}\|_\infty \leq 1$ by definition of $\lambda_{\max}$. Next, let V be the orthogonal complement of $\ker\mathbf{L} \subset \mathbb{R}^M$. A known property of the pseudoinverse operator [100, Corollary 7] is that $\mathbf{L}^T\mathbf{L}^{T\dagger}$ is the orthogonal projection operator onto V. By decomposing $\mathbf{v}(\mathbf{z}_{\text{opt}}) = \mathbf{v}_1 + \mathbf{v}_2$, where $\mathbf{v}_1 \in V$ and $\mathbf{v}_2 \in \ker\mathbf{L}$, we thus get $\mathbf{v}(\mathbf{z}_{\text{opt}}) + \lambda\mathbf{L}^T\mathbf{g} = \mathbf{v}_2$. Yet $\ker\mathbf{L} = \text{span}\{\mathbf{1}, \mathbf{x}\}$, since the canonical solutions $f_\mathbf{1}$ and $f_\mathbf{x}$ (that satisfy $f_\mathbf{1}(x_m) = 1$ and $f_\mathbf{x}(x_m) = x_m$ for every $m \in \{1, \dots, M\}$ respectively) are linear functions that are thus not penalized by the regularization. The optimality conditions of problem (39) (*i.e.*, setting the gradient to zero) then yield $\mathbf{v}(\mathbf{z}_{\text{opt}}) \perp \ker\mathbf{L}$, which implies that $\mathbf{v}_2 = \mathbf{0}$ and thus that $\mathbf{v}(\mathbf{z}_{\text{opt}}) + \lambda\mathbf{L}^T\mathbf{g} = \mathbf{0}$. This proves that $\mathbf{z}_{\text{opt}}$ satisfies the optimality condition of problem (36), and thus that $\mathbf{z}_{\text{opt}} = \mathbf{y}_\lambda = \beta_{\text{opt}_0}\mathbf{1} + \beta_{\text{opt}_1}\mathbf{x}$.

Item 3. Due to item 2, we have $\mathbf{y}_\lambda = \beta_{\text{opt}_0}\mathbf{1} + \beta_{\text{opt}_1}\mathbf{x}$ which implies that the points $\begin{bmatrix} x_m & y_{\lambda,m} \end{bmatrix}^T$ are aligned. Hence, the canonical dual certificate of the constrained problem (35) is $\eta_{\text{cano}} = 0$, which is nondegenerate. By Proposition 6, this implies that the unique solution to problem (35) is the canonical solution $f_{\mathbf{z}_{\text{opt}}} = f_{\max} = \beta_{\text{opt}_0} + \beta_{\text{opt}_1}(\cdot)$. Due to the equivalence between problems (35) and the (g-BLASSO) proved in Proposition 7, this concludes the proof.

References

[1] G. Cybenko, Approximation by superpositions of a sigmoidal function, Math. Control Signals Systems 2 (4) (1989) 303–314.
[2] K. Hornik, Approximation capabilities of multilayer feedforward networks, Neural Netw. 4 (2) (1991) 251–257.
[3] M. Leshno, V. Lin, A. Pinkus, S. Schocken, Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, Neural Netw. 6 (6) (1993) 861–867.

[4] T. Mitchell, Machine Learning, McGraw-Hill Education - Europe, 1997.
[5] Y. De Castro, F. Gamboa, Exact reconstruction using beurling minimal extrapolation, J. Math. Anal. Appl. 395 (1) (2012) 336–354.
[6] M. Unser, J. Fageot, J.P. Ward, Splines are universal solutions of linear inverse problems with generalized TV regularization, SIAM Rev. 59 (4) (2017) 769–793.
[7] H. Gupta, J. Fageot, M. Unser, Continuous-domain solutions of linear inverse problems with Tikhonov *vs.* generalized TV regularization, IEEE Trans. Signal Process. 66 (17) (2018) 4670–4684.
[8] C. Boyer, A. Chambolle, Y. De Castro, V. Duval, F. De Gournay, P. Weiss, On representer theorems and convex regularization, SIAM J. Optim. 29 (2) (2019) 1260–1281.
[9] C. Ekanadham, D. Tranchina, E. Simoncelli, Recovery of sparse translation-invariant signals with continuous basis pursuit, IEEE Trans. Signal Process. 59 (10) (2011) 4735–4744.
[10] V. Duval, G. Peyré, Sparse spikes super-resolution on thin grids II: the continuous basis pursuit, Inverse Problems 33 (9) (2017) 095008.
[11] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, SIAM Rev. 43 (1) (2001) 129–159.
[12] E. Candès, C. Fernandez-Granda, Towards a mathematical theory of super-resolution, Comm. Pure Appl. Math. 67 (6) (2014) 906–956.
[13] V. Duval, G. Peyré, Exact support recovery for sparse spikes deconvolution, Found. Comput. Math. 15 (5) (2015) 1315–1355.
[14] C. Fernandez-Granda, Super-resolution of point sources via convex programming, Inf. Inference: J. IMA (2016).
[15] S. Fisher, J. Jerome, Spline solutions to L^1 extremal problems in one and several variables, J. Approx. Theory 13 (1) (1975) 73–83.
[16] R. Koenker, P. Ng, S. Portnoy, Quantile smoothing splines, Biometrika 81 (4) (1994) 673–680.
[17] E. Mammen, S. van de Geer, Locally adaptive regression splines, Ann. Statist. 25 (1) (1997) 387–413.
[18] R. Tibshirani, Regression shrinkage and selection via the LASSO, J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1) (1996) 267–288.
[19] D. Donoho, Compressed sensing, IEEE Trans. Inform. Theory 52 (4) (2006) 1289–1306.
[20] E. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theory 52 (2) (2006) 489–509.
[21] Y. Eldar, G. Kutyniok, Compressed Sensing: Theory and Applications, Cambridge University Press, 2012.
[22] S. Foucart, H. Rauhut, A Mathematical Introduction to Compressive Sensing, vol. 1, Birkhäuser Basel, 2013.
[23] M. Unser, J. Fageot, H. Gupta, Representer theorems for sparsity-promoting ℓ_1 regularization, IEEE Trans. Inf. Theory 62 (9) (2016) 5167–5180.
[24] T. Hastie, R. Tibshirani, M. Wainwright, Statistical Learning with Sparsity: The LASSO and Generalizations, Chapman and Hall/CRC, 2015.
[25] I. Rish, G. Grabarnik, Sparse Modeling: Theory, Algorithms and Applications, CRC Press, 2014.
[26] D. Donoho, Superresolution via sparsity constraints, SIAM J. Math. Anal. (1992).
[27] B. Schölkopf, R. Herbrich, A. Smola, A generalized representer theorem, in: Computational Learning Theory, Springer, 2001, pp. 416–426.
[28] G. Wahba, Spline Models for Observational Data, SIAM, 1990.
[29] A. Berlinet, C. Thomas-Agnan, Reproducing Kernel Hilbert Spaces in Probability and Statistics, Springer Science & Business Media, 2011.
[30] A. Badoual, J. Fageot, M. Unser, Periodic splines and Gaussian processes for the resolution of linear inverse problems, IEEE Trans. Signal Process. 66 (22) (2018) 6047–6061.
[31] B. Adcock, A. Hansen, Generalized sampling and infinite-dimensional compressed sensing, Found. Comput. Math. 16 (5) (2016) 1263–1323.
[32] B. Adcock, A. Hansen, C. Poon, B. Roman, Breaking the coherence barrier: A new theory for compressed sensing, in: Forum of Mathematics, Sigma, vol. 5, Cambridge University Press, 2017.
[33] A. Bhandari, Y. Eldar, Sampling and super resolution of sparse signals beyond the Fourier domain, IEEE Trans. Signal Process. 67 (6) (2018) 1508–1521.
[34] B. Bodmann, A. Flinth, G. Kutyniok, Compressed sensing for analog signals, 2018, arXiv preprint arXiv:1803.04218.
[35] B. Bhaskar, G. Tang, B. Recht, Atomic norm denoising with applications to line spectral estimation, IEEE Trans. Signal Process. 61 (23) (2013) 5987–5999.
[36] K. Bredies, H. Pikkarainen, Inverse problems in spaces of measures, ESAIM Control Optim. Calc. Var. 19 (01) (2013) 190–218.
[37] S. Zuhovickiĭ, On approximation of real functions in the sense of P.L. Čebyšev, AMS Transl. Math. Monogr. 19 (2) (1962) 221–252.
[38] M. Krein, A. Nudelman, The Markov Moment Problem and Extremal Problems: Ideas and Problems of P.L. Cebysev and A.A. Markov and their Further Development, American Mathematical Society, 1977.
[39] E. Candès, C. Fernandez-Granda, Super-resolution from noisy data, J. Fourier Anal. Appl. (2013).
[40] J. Azais, Y. De Castro, F. Gamboa, Spike detection from inaccurate samplings, Appl. Comput. Harmon. Anal. (2015).
[41] B. Bhaskar, G. Tang, B. Recht, Near minimax line spectral estimation, IEEE Trans. Inform. Theory (2014).
[42] V. Duval, G. Peyré, Sparse regularization on thin grids I: the LASSO, Inverse Problems 33 (5) (2017) 055008.
[43] C. Poon, N. Keriven, G. Peyré, Support localization and the Fisher metric for off-the-grid sparse regularization, in: The 22nd International Conference on Artificial Intelligence and Statistics, 2019.
[44] Q. Denoyelle, V. Duval, G. Peyré, Support recovery for sparse super-resolution of positive measures, J. Fourier Anal. Appl. 23 (5) (2017) 1153–1194.
[45] C. Poon, G. Peyré, Multidimensional sparse super-resolution, SIAM J. Math. Anal. (2019).
[46] G. Schiebinger, E. Robeva, B. Recht, Superresolution without separation, Inf. Inference: J. IMA (2018).
[47] V. Duval, A characterization of the non-degenerate source condition in super-resolution, Inf. Inference: J. IMA (2019).
[48] H. García, C. Hernández, M. Junca, M. Velasco, Approximate super-resolution of positive measures in all dimensions, Appl. Comput. Harmon. Anal. (2020).
[49] Y. Chi, M. Ferreira Da Costa, Harnessing sparsity over the continuum: Atomic norm minimization for superresolution, IEEE Signal Process. Mag. 37 (2) (2020) 39–57.
[50] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci. 2 (1) (2009) 183–202.
[51] C. Elvira, R. Gribonval, C. Soussen, C. Herzet, OMP and continuous dictionaries: Is k-step recovery possible? in: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2019, pp. 5546–5550.
[52] Y. De Castro, F. Gamboa, D. Henrion, J.-B. Lasserre, Exact solutions to super resolution on semi-algebraic domains in higher dimensions, IEEE Trans. Inform. Theory 63 (1) (2016) 621–630.
[53] P. Catala, V. Duval, G. Peyré, A low-rank approach to off-the-grid sparse deconvolution, J. Phys. Conf. Ser. (2017) 012015.
[54] M. Frank, P. Wolfe, An algorithm for quadratic programming, Nav. Res. Logist. Q. (1956).
[55] N. Boyd, G. Schiebinger, B. Recht, The alternating descent conditional gradient method for sparse inverse problems, SIAM J. Optim. 27 (2) (2017) 616–639.
[56] Q. Denoyelle, V. Duval, G. Peyré, E. Soubies, The sliding Frank-Wolfe algorithm and its application to super-resolution microscopy, Inverse Problems (2019).
[57] J.-B. Courbot, V. Duval, B. Legras, Sparse analysis for mesoscale convective systems tracking, 2019, HAL archives ouvertes: hal-02010436ff.
[58] A. Flinth, F. de Gournay, P. Weiss, On the linear convergence rates of exchange and continuous methods for total variation minimization, Mathematical Programming 190 (1-2) (2020) 221–257.
[59] C. de Boor, On "best" interpolation, J. Approx. Theory 16 (1) (1976) 28–42.

[60] A. Pinkus, On smoothest interpolants, SIAM J. Math. Anal. 19 (6) (1988) 1431–1441.
[61] M. Unser, A representer theorem for deep neural networks, J. Mach. Learn. Res. 20 (110) (2019) 1–30.
[62] V. Duval, An epigraphical approach to the representer theorem, 2019, arXiv preprint arXiv:1912.13224.
[63] A. Flinth, P. Weiss, Exact solutions of infinite dimensional total-variation regularized problems, Inf. Inference: J. IMA 8 (3) (2019) 407–443.
[64] K. Bredies, M. Carioni, Sparsity of solutions for variational inverse problems with finite-dimensional data, Calc. Var. Partial Differential Equations 59 (1) (2019).
[65] M. Simeoni, Functional Inverse Problems on Spheres: Theory, Algorithms and Applications (Ph.D. thesis), Swiss Federal Institute of Technology Lausanne (EPFL), 2020.
[66] M. Simeoni, Functional penalised basis pursuit on spheres, Appl. Comput. Harmon. Anal. 53 (2021) 1–53.
[67] T. Debarre, S. Aziznejad, M. Unser, Hybrid-spline dictionaries for continuous-domain inverse problems, IEEE Trans. Signal Process. 67 (22) (2019) 5824–5836.
[68] J. Fageot, M. Simeoni, TV-based reconstruction of periodic functions, Inverse Problems 36 (11) (2020) 115015.
[69] M. Unser, J. Fageot, Native Banach spaces for splines and variational inverse problems, 2019, arXiv preprint arXiv:1904.10818.
[70] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.
[71] R. Pascanu, G. Montufar, Y. Bengio, On the number of response regions of deep feed forward networks with piece-wise linear activations, in: Second International Conference on Learning Representations (ICLR), ICLR, Banff, Canada, 2014.
[72] G. Montufar, R. Pascanu, K. Cho, Y. Bengio, On the number of linear regions of deep neural networks, in: Advances in Neural Information Processing Systems, vol. 27, Curran Associates, Inc., 2014, pp. 2924–2932.
[73] R. Balestriero, R. Baraniuk, Mad max: affine spline insights into deep learning, Proceedings of the IEEE 109 (5) (2021) 704–727.
[74] F. Agostinelli, M. Hoffman, P. Sadowski, P. Baldi, Learning activation functions to improve deep neural networks, in: Third International Conference on Learning Representations (ICLR) Workshop, ICLR, San Diego, USA, 2015.
[75] S. Aziznejad, H. Gupta, J. Campos, M. Unser, Deep neural networks with trainable activations and controlled lipschitz constant, IEEE Transactions on Signal Processing 68 (2020) 4688–4699.
[76] I. Daubechies, R. DeVore, S. Foucart, B. Hanin, G. Petrova, Nonlinear approximation and (deep) relu networks, Constructive Approximation (2021).
[77] T. Poggio, L. Rosasco, A. Shashua, N. Cohen, F. Anselmi, Notes on hierarchical splines, DCLNs and i-theory, in: Cbmm Memo, Center for Brains, Minds and Machines (CBMM), 2015.
[78] H. Bölcskei, P. Grohs, G. Kutyniok, P. Petersen, Optimal approximation with sparsely connected deep neural networks, SIAM J. Math. Data Sci. 1 (1) (2019) 8–45.
[79] D. Yarotsky, Error bounds for approximations with deep ReLU networks, Neural Netw. 94 (2017) 103–114.
[80] P. Petersen, F. Voigtlaender, Optimal approximation of piecewise smooth functions using deep ReLU neural networks, Neural Netw. 108 (2018) 296–330.
[81] R. Gribonval, G. Kutyniok, M. Nielsen, F. Voigtlaender, Approximation spaces of deep neural networks, 2019, working paper or preprint.
[82] J. de Dios, J. Bruna, On sparsity in overparametrised shallow ReLU networks, 2020, arXiv preprint arXiv:2006.10225.
[83] I. Savarese, D. Soudry, N. Srebro, How do infinite width bounded norm networks look in function space? in: A. Beygelzimer, D. Hsu (Eds.), Proceedings of the Thirty-Second Conference on Learning Theory, in: Proceedings of Machine Learning Research, vol. 99, PMLR, Phoenix, USA, 2019, pp. 2667–2690.
[84] R. Parhi, R. Nowak, The role of neural network activation functions, IEEE Signal Process. Lett. 27 (2020) 1779–1783.
[85] R. Parhi, R. Nowak, Banach space representer theorems for neural networks and ridge splines, J. Mach. Learn. Res. 22 (43) (2021) 1–40.
[86] G. Ongie, R. Willett, D. Soudry, N. Srebro, A function space view of bounded norm infinite width ReLU nets: The multivariate case, in: International Conference on Learning Representations, 2020.
[87] L. Schwartz, Théorie des Distributions, Hermann, 1966.
[88] T. Debarre, J. Fageot, H. Gupta, M. Unser, B-spline-based exact discretization of continuous-domain inverse problems with generalized TV regularization, IEEE Trans. Inform. Theory (2019).
[89] R. Tibshirani, The LASSO problem and uniqueness, Electron. J. Stat. 7 (2013) 1456–1490.
[90] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends® Mach. Learn. 3 (1) (2010) 1–122.
[91] B. He, X. Yuan, On the $O(1/n)$ convergence rate of the douglas-rachford alternating direction method, SIAM J. Numer. Anal. 50 (2) (2012) 700–709.
[92] W. Deng, W. Yin, On the global and linear convergence of the generalized alternating direction method of multipliers, J. Sci. Comput. 66 (3) (2015) 889–916.
[93] F. Bach, R. Jenatton, J. Mairal, G. Obozinski, Optimization with sparsity-inducing penalties, Found Trends® Mach. Learn. 4 (1) (2011) 1–106.
[94] M.R. Osborne, B. Presnell, B.A. Turlach, On the LASSO and its dual, J. Comput. Graph. Statist. 9 (2) (2000) 319.
[95] E. Soubies, F. Soulez, M. McCann, T.-a. Pham, L. Donati, T. Debarre, D. Sage, M. Unser, Pocket guide to solve inverse problems with GlobalBioIm, Inverse Problems 35 (10) (2019) 1–20, paper no. 104006.
[96] M. Reed, B. Simon, Methods of Modern Mathematical Physics: Functional Analysis, vol. 1, Gulf Professional Publishing, 1980.
[97] I. Ekeland, R. Temam, Convex Analysis and Variational Problems, SIAM, 1976.
[98] W. Rudin, Real and Complex Analysis, McGraw-Hill Education, 1986.
[99] R. Rockafellar, Convex Analysis, Princeton University Press, 1970.
[100] A. Ben-Israel, T. Greville, Generalized Inverses, Springer-Verlag, 2003.