Exact solutions of infinite dimensional total-variation regularized problems

Axel Flinth, Pierre Weiss

Exact solutions of infinite dimensional total-variation regularized problems.

Axel Flinth & Pierre Weiss

October 31, 2017

Abstract

We study the solutions of infinite dimensional linear inverse problems over Banach spaces. The regularizer is defined as the total variation of a linear mapping of the function to recover, while the data fitting term is a near arbitrary function. The first contribution describes the solution's structure: we show that under mild assumptions, there always exists an m-sparse solution, where m is the number of linear measurements of the signal. Our second contribution is about the computation of the solution. While most existing works first discretize the problem, we show that exact solutions of the infinite dimensional problem can be obtained by solving one or two consecutive finite dimensional convex programs depending on the measurement functions structures. These results extend recent advances in the understanding of total-variation regularized inverse problems.

1 Introduction

Let $u \in \mathcal{B}$ be a signal in some vector space $\mathcal{B}$ and assume that it is probed indirectly, with m corrupted linear measurements:

$$b = P(Au),$$

where $A : \mathcal{B} \to \mathbb{R}^m$ is a measurement operator defined by $(Au)_i = \langle a_i, u \rangle$, each a_i being an element in $\mathcal{B}^*$, the dual of $\mathcal{B}$. The mapping $P : \mathbb{R}^m \to \mathbb{R}^m$ denotes a perturbation of the measurements, such as quantization, additional Gaussian or Poisson noise, or any other common degradation operator. Inverse problems consist in estimating u from the measurements b. Assuming that $\dim(\mathcal{B}) > m$, it is clearly impossible to recover u knowing b only. Hence, various regularization techniques have been proposed to stabilize the recovery.

Probably the most well known and used example is Tikhonov regularization [21], which consists in minimizing quadratic cost functions. The regularizers are particularily appreciated for their ease of analysis and implementation. Over the last 20 years, sparsity promoting regularizers have proved increasingly useful, especially when the signals to recover have some underlying sparsity structure. Sparse regularization can be divided into two categories: the analysis formulation and the synthesis formulation.

The analysis formulation consists in solving optimization problems of the form

$$\inf_{u \in \mathcal{B}} J(u) := f_b(Au) + \|Lu\|_{TV}, \tag{1}$$

where $f_b : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is an application dependent data fidelity term and $L : \mathcal{B} \to \mathcal{E}$ is a linear operator, mapping $\mathcal{B}$ to some space $\mathcal{E}$ such as $\mathbb{R}^n$, the space of sequences in ℓ_1 or the space of Radon measures $\mathcal{M}$. The total variation norm $\|\cdot\|_{TV}$ coincides with the ℓ_1-norm when $\mathcal{E}$ is discrete, but it is more general since it also applies to measures.

The synthesis formulation on its side consists in minimizing

$$\inf_{\mu \in \mathcal{E}} f_b(AD\mu) + \|\mu\|_{TV}, \tag{2}$$

where $D : \mathcal{E} \to \mathcal{B}$ is the linear synthesis operator, also called dictionary. The estimate of u in that case reads $\hat{u} = D\hat{\mu}$, where $\hat{\mu}$ is a solution of (2).

Problems (1) and (2) triggered a massive interest from both theoretical and practical perspectives. Among the most impressive theoretical results, one can cite the field of compressed sensing [9] or super-resolution [8, 15], which certify that under suitable assumptions, the minimizers of (1) or (2) coincide with the true signal u.

Most of the studies in this field are confined to the case where both $\mathcal{B}$ and $\mathcal{E}$ are finite dimensional [9, 13, 17, 19]. In the last few years, some efforts have been provided to get a better understanding of (1) and (2) where $\mathcal{B}$ and $\mathcal{E}$ are sequence spaces [2, 3, 29, 28]. Finally, a different route, which will be followed in this paper, is the case where $\mathcal{E} = \mathcal{M}$, the space of Radon measures on a continuous domain. In that case, problems (1) and (2) are infinite dimensional problems over measure spaces. One instance in that class is that of total variation minimization (in the PDE sense [4], that is the total variation of the distributional derivative), which became extremely popular in the field of imaging since its introduction in [24]. There has been surge of interest in understanding the fine properties of the solutions in this setting, with many significant results [7, 8, 26, 15, 10, 30]. The aim of this paper is to continue these efforts by bringing new insights in a general setting.

Contributions and related works The main contributions are twofold: one is about the structure of the solutions of (1), while the other is about how to numerically solve this problem without discretization. The results directly apply to problem (2) since, with regards to our concerns, the synthesis problem (2) is a special case of the analysis problem (1). It indeed suffices to take $L = \mathrm{Id}$ and $\mathcal{B} = \mathcal{M}$ for (2) to be an instance of (1). Notice however that in general, the two approaches should be studied separately [17].

On the theoretical side, we provide a theorem characterizing the structure of the solutions of problem (1) under certain assumptions on the operator L. Roughly speaking, this theorem states that there always exist m-sparse solutions. The precise meaning of this claim will be clarified in Theorem 1. This result is strongly related and was actually motivated by [30]. In there, the authors restrict their study to certain stationary operators L over spaces of functions defined on $\Omega = \mathbb{R}^d$. Their main result states that in that case, generalized splines with m knots actually describe the whole set of solutions. Similar

results [18] were actually obtained much earlier on bounded domains and seem to have remained widely ignored until they were revitalised by Unser-Fageot-Ward. The value of our result lies in the fact that it holds for more general classes of operators L, spaces $\mathcal{B}$, domains Ω and functions f_b. Furthermore, the proof technique is different from [30]: it is constructive and presumably applicable to wider settings.

On the numerical side, let us first emphasize that in an overwhelming number of works, problem (1) is solved by first discretizing the problem to make it finite dimensional and then approximate solutions are found with standard procedures from convex programming. Theories such as Γ-convergence [6] then sometimes allow showing that as the discretization parameter goes to 0, solutions of the discretized problem converge (in a weak sense) to the solutions of the continuous problem. In this paper, we show that under some assumptions on the measurement functions (a_i), the infinite dimensional problem (1) can be attacked directly without discretization: the resolution of one or two consecutive finite dimensional convex programs allows recovering *exact solutions* to problem (1) or (2). The structure of the convex programs depend on the structure of measurement vectors. Once again, this result is strongly related to recent advances. For instance, it is shown in [8, 26] that a specific instance of (2) with $L = \mathrm{Id}$ can be solved exactly thanks to semi-definite relaxation or Prony type methods when the signal domain is the torus $\Omega = \mathbb{T}$ and the functions (a_i) are trigonometric polynomials. Similar results were obtained in [12] for more general semi-algebraic domains using Lasserre hierarchies [22]. Once again, the value of our paper lies in the fact that it holds for near arbitrary convex functions f_b and for a large class of operators L such as the derivative. To the best of our knowledge, the only case considered until now was $L = \mathrm{Id}$. In addition, our results provide some insight on the standard minimization strategy: we show that it corresponds to solving a different infinite dimensional problem *exactly*, where the sampling functions are piecewise linear. We also show that the solution of the standard discretization can be made sparser by merging Dirac masses located on neighboring grid points.

2 Main results

2.1 Notation

In all of the paper, $\Omega \subseteq \mathbb{R}^d$ denotes an open subset either bounded or unbounded. The space of distributions on Ω is denoted $\mathcal{D}^*(\Omega)$. We let $\mathcal{M}(\Omega)$ denote the set of Radon measures on Ω, i.e. the dual $\mathcal{C}_0(\Omega)^*$ of $\mathcal{C}_0(\Omega)$, the space of continuous functions on Ω vanishing at infinity:

$$\mathcal{C}_0(\Omega) \quad = \quad \left\{ \begin{array}{ll} f : \Omega \to \mathbb{R}, & f \text{ continuous}, \\ & \forall \epsilon > 0, \exists C \subset \Omega \text{ compact }, \forall x \notin C, |f(x)| < \epsilon \end{array} \right\}.$$

We will throughout the whole paper view $(\mathcal{M}(\Omega), \|\cdot\|_{TV})$ as a Banach space, and not, as often is done, as a locally convex space equipped with the weak-$*$-topology. When we do this, $\mathcal{C}_0(\Omega)$ is a subset, and not the whole of, the dual $\mathcal{M}^*$ of $\mathcal{M}$ (as it would have been if we have viewed $\mathcal{M}$ as a locally convex space).

Let $J : \mathcal{B} \to \mathbb{R} \cup \{+\infty\}$ denote a convex lower-semicontinuous function. We let J^* denote its Fenchel conjugate and $\partial J(u)$ denote its subdifferential at

$u \in \mathcal{B}$. Let $X \subset \mathcal{E}$ be a subset of some vector space $\mathcal{E}$. The indicator function of X is defined for all $e \in \mathcal{E}$ by:

$$\iota_X(e) = \begin{cases} 0 & \text{if } e \in X \\ +\infty & \text{otherwise.} \end{cases}$$

We refer the reader to [16] for more insight on convex analysis in vector spaces.

Remark 1. *All the results in our paper hold when Ω is a separable, locally compact topological space such as the torus $\mathbb{T} = \mathbb{R}\backslash\mathbb{N}$. The proofs require minor technical amendments related to the way the space is discretized. We chose to keep a simpler presentation in this paper.*

2.2 Assumptions

Let us describe the setting in which we will prove the main result in some detail. Let $L : \mathcal{D}'(\Omega) \to \mathcal{M}(\Omega)$ be a continuous linear operator defined on the space of distributions $\mathcal{D}'(\Omega)$. Consider the following linear subspace of $\mathcal{D}^*(\Omega)$

$$\mathcal{B}^\circ = \{u \in \mathcal{D}'(\Omega) \,|\, Lu \in \mathcal{M}(\Omega)\}\,.$$

Now, let $\|\cdot\|_K$ be a semi-norm on $\mathcal{B}^\circ$, which restricted to $\ker L$ is a norm. We define

$$\mathcal{B} = \{u \in \mathcal{D}'(\Omega) \,|\, Lu \in \mathcal{M}(\Omega), \|u\|_K < \infty\}\,,$$

and equip it with the norm $\|u\| = \|Lu\|_{TV} + \|u\|_K$. We will assume that

Assumption 1 (Assumption on $\mathcal{B}$)**.** *$\mathcal{B}$ is a Banach space.*

We will make the following additional structurial assumptions on the map L:

Assumption 2 (Assumptions on L)**.**

- *The kernel of L has a complementary subspace, i.e. a closed subspace V such that* $\ker L \oplus V = \mathcal{B}$.
- *The range of L is closed, and has a complementary subspace W, i.e.,* $\operatorname{ran} L \oplus W = \mathcal{M}$.

An important special case of operators satisfying the assumption 2 are *Fredholm operators* for which the space W complementary to $\operatorname{ran} L$ is finite-dimensional, and $\ker L$ is itself finite dimensional, see e.g. [25, Lemma 4.21].

The restriction $L|_V$ of $L : V \to \operatorname{ran} L$ is a bijective operator, and therefore has a continuous inverse $(L|_V)^{-1}$, by the continuous inverse theorem. With the help of this inverse, we can define a *pseudoinverse* $L^+ : \mathcal{M} \to \mathcal{B}$ through

$$L^+ = j_V (L|_V)^{-1} \Pi_{\operatorname{ran} L},$$

where j_V denotes the injection $V \hookrightarrow \mathcal{B}$ and $\Pi_{\operatorname{ran} L}$ the projection from M to $\operatorname{ran} L$. Both of these as well as $(L|_V)^{-1}$ are continuous, so that L^+ is continuous.

We will furthermore have to restrict the functionals a_i used to probe the signals slightly.

Assumption 3 (Assumption on a_i)**.**

The functionals $a_i \in \mathcal{B}^*(\Omega)$ have the property that $(L^+)^* a_i \in \mathcal{C}_0(\Omega)$. That is, there exist functions $\rho_i \in \mathcal{C}_0(\Omega)$ with

$$\forall \mu \in \mathcal{M} : \left\langle (L^+)^* a_i, \mu \right\rangle = \int_\Omega \rho_i(x) d\mu(x).$$

This assumption may seem a bit artificial, but we will see that it is crucial, both in the more theoretical first part of the paper, as well as in the second one dealing with the numerical resolution of the problem. Furthermore, it is equivalent to an assumption in the main result of [30], as will be made explicit in the sequel.

Until now, we have not touched upon the properties of the function f_b. We do this implicitly with the following condition:

Assumption 4 (Solvability Assumption)**.** *The problem* (1) *has at least one solution.*

This assumption is of course necessary in order to make questions about the structure of the solutions of (1) to make sense at all. A myriad of problems have this property, as the following simple proposition shows:

Proposition 1. *Assume that f_b is lower semi-continuous and coercive (i.e. $\lim_{\|x\|_2 \to \infty} f(x) = \infty$), and that $f_b \circ A$ has a non-empty domain. Then, under assumptions 1, 2 and 3, the problem* (1) *has a solution.*

The proof, which relies on standard arguments, can be found in Section 4.4. Let us here instead argue that the assumptions in 1 are quite light and cover many common data fidelity terms as exemplified below.

Equality constraints This case corresponds to

$$f_b(x) = \iota_{\{b\}}(x) = \begin{cases} 0 & \text{if } x = b \\ +\infty & \text{otherwise.} \end{cases} \tag{3}$$

This data fidelity term is commonly used when the data is not corrupted. A solution exists if $b \in \operatorname{ran}(A)$. The two super-resolution papers [8, 26] use this assumption.

Quadratic The case $f_b(x) = \frac{\lambda}{2}\|C^{-1}(x-b)\|_2^2$, where $\lambda > 0$ is a data fit parameter, is commonly used when the data suffers from additive Gaussian noise with a covariance matrix C.

ℓ_1-norm When data suffers from outliers, it is common to set $f_b(x) = \lambda\|x-b\|_1$, with $\lambda > 0$.

Box constraints When the data is quantized, a natural data fidelity term is a box constraint of the following type

$$f_b(x) = \begin{cases} 0 & \text{if } \|C(x-b)\|_\infty \leq 1 \\ +\infty & \text{otherwise,} \end{cases}$$

where $C \in \mathbb{R}^{m \times m}$ is a diagonal matrix with positive entries.

Phase Retrieval Many non-convex functions f_b fulfill our assumptions. In particular, any of the above fidelity terms can be combined with the (pointwise) absolute value to yield a feasible function f_b, i.e. for instance

$$f_b(x) = \iota_{\{b\}}(|x|) = \begin{cases} 0 & \text{if } |x| = b \\ +\infty & \text{otherwise.} \end{cases}.$$

Such functions appear in the *phase retrieval problem*, where one tries to reconstruct a signal u from absolute values of type $|Au|$.

2.3 Structure of the solutions

We are now ready to state the first important result of this paper.

Theorem 1. *Under assumptions 1, 2, 3 and 4, problem* (1) *has a solution of the form*

$$\hat{u} = u_K + \sum_{k=1}^{p} d_k L^+ \delta_{x_k},$$

with $p \leq \overline{m} = m - \dim(A^*(\ker L))$, $u_K \in \ker L$, $d = (d_k)_{1 \leq k \leq p}$ *in* $\mathbb{R}^p$ *and* $X = (x_k)_{1 \leq k \leq p}$ *in* Ω^p.

The proof of this theorem consists of three main steps. We provide the first two below, since they are elementary and provide some insight on the theorem. The last step appears in many works. We provide an original proof in the appendix.

Proof.
Step 1: In this step, we transform the data fitting f_b into an equality constraint. To see why this is possible, let $\overline{u}$ be a solution of the problem (1). Then any solution of the problem

$$\min_{u \in \mathcal{B}} \|Lu\|_{TV} \text{ subject to } Au = A\overline{u} =: y$$

will also be a solution $\hat{u}$ of (1), since it satisfies $f_b(A\hat{u}) = f_b(A\overline{u})$ and $\|L\overline{u}\|_{TV} = \|L\hat{u}\|_{TV}$. Those two equalities are required, otherwise, $\overline{u}$ would not be a solution since $J(\hat{u}) < J(\overline{u})$.

Step 2: In this step, we show that it is possible to discard the operator L. To see this, notice that since every $u \in \mathcal{B}$ can be written as $L^+\mu + u_K$ with $\mu \in \mathcal{M}$ and $u_K \in \ker L$. Therefore, we have

$$\begin{aligned} &\left(\min_{u \in \mathcal{B}} \|Lu\|_{TV} \text{ subject to } Au = y \right) \\ &\quad = \left(\min_{u_K \in \ker(L), \mu \in \mathcal{M}} \|\mu\|_{TV} \text{ subject to } A(u_K + L^+\mu) = y \right) \end{aligned}$$

Now, set $X = A\ker(L)$. Since X is a finite-dimensional subspace of $\mathbb{R}^m$, we may decompose $y = y_X + y_{X^\perp}$, with $y_X \in X$ and $y_{X^\perp} \in X^\perp$, the orthogonal complement of X in $\mathbb{R}^m$. Notice that for every $\mu \in \mathcal{M}$, there exists a $u_K \in \ker L$

with $A(u_K + L^+\mu) = y$ if and only if $\Pi_{X^\perp} AL^+\mu = y_{X^\perp}$. Hence, the above problems can be simplified as follows

$$\min \|\mu\|_{TV} \text{ subject to } H\mu = y_{X^\perp}, \tag{4}$$

with $H : \mathcal{M} \to X^\perp$, $H = \Pi_{X^\perp} A^* L^+$, with $\dim X^\perp = \overline{m}$.

Step 3: The last step consists in proving that the problem (4) has a solution of the form $\sum_{k=1}^{p} d_k \delta_{x_k}$. This result is well-known when Ω is a countable set, see e.g. [29]. It is also available in infinite dimensions on compact domains. We refer to [18] for instance, for an early proof, based on the Krein-Milmann theorem. We propose an alternative strategy in the appendix based on a discretization procedure. □

Remark 2. *In [18, 30], the authors further show that the extremal points of the solution set are of the form given in Theorem 1, if f_b is the indicator function of a closed convex set. Their argument is based on a proof by contradiction. Following this approach, it is possible to prove the same result in our setting. We choose not to carry out the details about this since we also wish to cover nonconvex problems.*

Before going further, let us show some consequences of this theorem.

2.3.1 Example 1: $L = \mathrm{Id}$ and the space $\mathcal{M}$

Probably the easiest case consists in choosing an arbitrary open domain $\Omega \subseteq \mathbb{R}^d$, to set $\mathcal{B} = \mathcal{M}(\Omega)$ and $L = \mathrm{Id}$. In this case, all the assumptions 2 on L are trivially met. We have $\operatorname{ran} \mathrm{Id} = \mathcal{M}(\Omega)$, $\ker \mathrm{Id} = \{0\}$ and $\mathrm{Id}^+ = \mathrm{Id}$. Therefore, Theorem 1 in this specific case guarantees the existence of a minimizer of (1) of the form

$$\hat{\mu} = \sum_{k=1}^{p} d_k \delta_{x_k},$$

with $p \leq m$. The assumption 3 in this case simply means that the functionals a_i can be identified with continuous operators vanishing at infinity.

Note that the synthesis formulation (2) can be seen as a subcase of this setting. The structure of the minimizing measure in Theorem 1 implies that the signal estimate $\hat{u}$ has the following form

$$\hat{u} = D\hat{\mu} = \sum_{k=1}^{p} d_k D\delta_{x_k}.$$

The vectors $(D\delta_x)_{x\in\Omega}$ can naturally be interpreted as the atoms of a dictionary. Hence, Theorem 1 states that there will always exist at least one estimate from the synthesis formulation which is sparsely representable in the dictionary $(D\delta_x)_{x\in\Omega}$.

2.3.2 Example 2: Spline-admissible operators and their native spaces

The authors of [30] consider a generic operator L defined on the space of tempered distributions $\mathcal{S}'(\mathbb{R}^d)$ and mapping into $\mathcal{M}(\mathbb{R}^d)$, which is

- Shift-invariant,
- for which there exists a function ρ_L (a generalized spline) of polynomial growth, say

$$\operatorname{esssup}_{x\in\mathbb{R}^d} |\rho_L(x)| \, (1 + \|x\|)^{-r} < +\infty, \tag{5}$$

  obeying $L\rho_L = \delta_0$.
- The space of functions in the kernel of L obeying the growth estimate (5) is finite dimensional.

The authors call such operators *spline-admissible*. A typical example is the distributional derivative D on $\Omega = \mathbb{R}$. For each such operator L, they define a space $\mathcal{M}_L(\mathbb{R}^d)$ as the set of functions f obeying the growth estimate (5) while still having the property $Lf \in \mathcal{M}(\mathbb{R}^d)$. The norm on $\mathcal{M}_L$ is as in our formulation, whereby $\|\cdot\|_K$ is defined through a dual basis of a (finite) basis of $\ker L$.

They go on to prove that $\mathcal{M}_L(\mathbb{R}^d)$ is a Banach space, which has a separable predual $\mathcal{C}_L(\mathbb{R}^d)$, and (in our notation) assume that the functionals $a_i \in M_L^*(\mathbb{R}^d)$ can be identified with elements of $\mathcal{C}_L(\mathbb{R}^d)$.

It turns out that using this construction, the operator L and functionals (a_i) obey the assumptions 2 and 3, respectively.

Proposition 2.

- *The operator $L : \mathcal{M}_L(\mathbb{R}^d) \to \mathcal{M}(\mathbb{R}^d)$ is Fredholm. In fact,* $\operatorname{ran} L$ *is even equal to $\mathcal{M}(\mathbb{R}^d)$.*
- *The functionals $a_i \in M_L^*(\mathbb{R}^d)$ obey assumption 3. In fact, we even have*

$$(L^+)^* a \in \mathcal{C}_0(\mathbb{R}^d) \iff a \in \mathcal{C}_L(\mathbb{R}^d).$$

Hence, the assumptions in [30] are a special case of the ones used in this paper.

2.3.3 Example 3: More general differential operators and associated spaces

The inclusion of operators with infinite dimensional kernel allows us to treat differential operators in a bit more streamlined way than above, in particular removing the restricted growth conditions. Let Ω be an open subset of $\mathbb{R}^d$ and $P(D)$ a differential operator on Ω, i.e. an expression of the form

$$P(D) = \sum_{|\alpha|\leq K} p_\alpha D^\alpha,$$

where $D^\alpha = \partial_{\alpha_1} \cdot \partial_{\alpha_j}$ is a partial derivative operator and p_α are measurable functions on Ω. Note that $P(D)$ does not need to be shift invariant (if $\Omega \neq \mathbb{R}^d$, shift-invariance is not even possible to define).

In order to define the norm of functions in the kernel of $L = P(D)$ properly, which we will not assume to satisfy any growth conditions, we assume that there exists a bounded subset $K \subseteq \Omega$ with the following *continuation property:*

Assumption 5 (Continuation property)**.** *For each distribution $u \in \mathcal{D}'(K)$ with $P(D)u = 0$, there exists exactly one $\widehat{u} \in \mathcal{D}'(\Omega)$ with $P(D)\widehat{u} = 0$ in Ω and $\widehat{u} = u$ in K.*

We will see that for a large class of elliptic operators, we can choose K to be any bounded set with non-empty interior and smooth boundary. These conditions will furthermore in particular prove that $\|\cdot\|_{\mathcal{M}(K)}$ is a seminorm on a space $\mathcal{B}$, which restricted to $\ker P(D)$ is a norm.

The fundamental assumption we will make is the following:

Assumption 6 (Green function hypothesis)**.** *For each $x \in \Omega$, there exists a solution $u_x \in \mathcal{C}(\Omega)$ of the problem*

$$P(D)u_x = \delta_x. \tag{6}$$

We also assume that the map $\Omega \ni x \mapsto u_x \in \mathcal{C}(\Omega)$ is continuous and bounded, i.e. $\sup_{x\in\Omega} \|u_x\|_\infty < \infty$.

Now we define, inspired by the native spaces $\mathcal{M}_L$ from above, a space $\mathcal{B}_P$, which $P(D)$ naturally sends to $\mathcal{M}(\Omega)$:

$$\mathcal{B}_P = \{u \in \mathcal{D}'(\Omega) \,|\, P(D)u \in \mathcal{M}(\Omega), u|_K \in \mathcal{M}(K)\}.$$

Lemma 1. *Under assumptions 5 and 6, the following holds: The expression*

$$\|u\|_{\mathcal{B}_P} = \|P(D)u\|_{TV} + \|u|_K\|_{TV}$$

defines a norm on $\mathcal{B}_P$. $\mathcal{B}_P$ equipped with this norm is a Banach space, i.e., satisfies assumption 1.

We now prove that 6 implies that $P(D)$ obeys the assumption 2, and state a more specific one which implies that relatively general functionals a obey assumption 3. To simplify the formulation of it slightly, let us introduce the following notion: we say that a mapping $T : \Omega \to \mathcal{C}(\Omega)$ vanishes at infinity on compact sets if for each compact subset $C \subseteq \Omega$, the function $x \mapsto \sup_{y\in C} |T(x)(y)|$ vanishes at infinity.

Proposition 3. *Under assumption 6, $L = P(D)$ satisfies assumptions 2. In particular, $\operatorname{ran} L = \mathcal{M}$, and the operator L^+ is given by*

$$(L^+\mu)(x) = \int_\Omega u_y(x) d\mu(y) \tag{7}$$

Furthermore, if a is a functional of the type

$$\langle a, u\rangle = \int_\Omega a(x)u(x)dx, \tag{8}$$

with $a \in L^1(\Omega)$, we have

$$((L^+)^* a)(x) = \int_\Omega u_x(y)a_i(y)dy$$

$(L^+)^*a$ obeys the assumption 3 provided the map $x \mapsto u_x$ vanishes at infinity on compact sets.

Remark 3. *Since we have assumed no growth restriction on the elements of* $\ker P(D)$*, in general, not every function* $a \in L^1(\Omega)$ *will cause* (8) *to define a functional on* $\mathcal{B}_P$*. However,* if *this is the case for a* specific *a,* $(L^+)^* a$ *will be well-defined and have the claimed form.*

An example of an additional assumption which will make (8) *actually define a functional on* $\mathcal{B}_P$ *is that a is continuous and has compact support inside* K*, since then*

$$|\langle a, u \rangle| \leq \|a\|_\infty \|u|_K\|_{TV} \leq \|a\|_\infty \|u\|_{\mathcal{B}_P}.$$

Let us now give a relatively general example of operators which satisfy the properties presented in Proposition 3. It for instance includes poly-Laplacian operators Δ^k of sufficiently high order on for $\Omega \subseteq \mathbb{R}^d$, either bounded or equal to the entire space $\mathbb{R}^n$.

Let $k \in \mathbb{N}$ and $P(D)$ be a differential operator on Ω of the form

$$P(D) = \sum_{|\alpha|=k} \sum_{|\beta|=k} D^\beta (p_{\alpha,\beta}(x) D^\alpha) \tag{9}$$

for some bounded functions $p_{\alpha,\beta} \in \mathcal{C}^k(\Omega)$. Also assume that $P(D)$ obeys the following ellipticity condition

$$\inf_{x \in \Omega} \inf_{\|\xi\|_2 = 1} \sum_{|\alpha|=k} \sum_{|\beta|=k} p_{\alpha,\beta}(x) \xi^{\alpha+\beta} =: C > 0. \tag{10}$$

Proposition 4. *Suppose that* $k > \frac{d}{2}$*. For either* Ω *bounded with Lipschitz domain or* $\Omega = \mathbb{R}^d$*, the following is true. Under the ellipticity assumption* (10)*, the problem* (6) *admits for each* $x \in \Omega$ *a solution* $u_x \in \mathcal{C}(\Omega)$*. The map* $x \to u_x$ *is furthermore vanishing at infinity on compact sets.*

Also, any set K *with non-empty interior and smooth domain obeys assumption 5.*

Remark 4. *The assumption* $k > d/2$ *is crucial, since only then, we can guarantee that the solutions* u_x *of* (6) *are continuous. Consider for instance the Laplacian operator* Δ *on* $\mathbb{R}^d$ *for* $d \geq 2$*. Then* $k = 1 \leq d/2$ *and*

$$u_x(y) = \begin{cases} \log(\|y - x\|_2) & \text{if } d = 2, \\ \|x - y\|_2^{2-d} & \text{otherwise,} \end{cases}$$

which are not continuous.

2.3.4 Example 4: $L = D$ and the space $BV(]0,1[)$

Another important operator which is *not* covered by Proposition 3 is the univariate derivative D in the univariate case. In this case, the function u_x in (6) is equal to a shifted Heaviside function, which of course is not continuous. At least this operator can however still be naturally included in our framework, as we will show here.

We set $\Omega =]0,1[$. The space $BV(\Omega)$ of bounded variation functions is defined by (see [4]):

$$BV(\Omega) = \{u \in L^1(\Omega), Du \text{ is a Radon measure}, \|Du\|_{TV} < +\infty\}, \tag{11}$$

where D is the distributional derivative. Using our notations, it amounts to taking $L = D$ and $\mathcal{B} = BV(\Omega)$. For this space, we have $\ker L = \text{span}(1)$, the vector space of constant functions on Ω. (Note that in fact, the norm $\|u\|_{BV} = \|u\|_1 + \|Du\|_{TV}$ is of the general form described in the introduction).

Lemma 2. *For $L = D$ we have $\operatorname{ran} L = \mathcal{M}$, and for all $\mu \in \mathcal{M}$ and all $s \in [0, 1]$,*

$$(L^+\mu)(s) = \mu([0, s]) - \int_0^1 \mu([0, t])\, dt. \tag{12}$$

In addition, for a functional $\xi \in BV(]0, 1[)^*$ of the form

$$\langle \xi, u \rangle = \int_0^1 \xi(t)u(t)\, dt,$$

with $\xi \in L^1(\Omega)$, we have $(L^+)^*\xi \in \mathcal{C}_0(\Omega)$ and letting $\bar{\xi} = \int_0^1 \xi(t)\, dt$, we have

$$((L^+)^*\xi)(s) = \int_0^s (\bar{\xi} - \xi(t))\, dt. \tag{13}$$

As can be seen, L^+ is simply a primitive operator. The elementary functions $L^+\delta_x$ are Heavyside functions translated at a distance x from the origin. Hence, Theorem 1 states that there always exist total variation minimizers in 1D that can be written as staircase functions with at most $\overline{m}$ jumps. Note that in this case, the Heavyside functions coincide with the general splines introduced in [30].

2.3.5 An uncovered case: $L = \nabla$ and the space $BV(]0, 1[^2)$

It is very tempting to use Theorem 1 on the space $\mathcal{B} = BV(]0, 1[^2)$. As mentioned in the introduction, this space is crucial in image processing since its introduction in [24]. Unfortunately, this case is not covered by Theorem 1, since $L\mathcal{B}$ is then a space of vector valued Radon measures, and our assumptions only cover the case of scalar measures.

2.4 Numerical resolution

In this section, we show how the infinite dimensional problem (1) can be solved using standard optimization approaches. We will make the following additional assumption:

Assumption 7 (Additional assumption on f_b)**.** *f_b is convex and lower semi-continuous.*

Depending on the structure of the measurement functions (a_i), we will propose to solve the primal problem (1) directly, or to solve two consecutive convex problems: the dual and the primal. We first recollect a few properties of the dual to shed some light on the solutions properties.

2.4.1 The dual problem and its relationship to the primal

A natural way to turn (1) into a finite dimensional problem is to use duality as shown in the following proposition.

Proposition 5 (Dual of problem (1))**.** *Define* $h : \mathcal{M}(\Omega) \to \mathbb{R} \cup \{+\infty\}$ *by* $h(\mu) = \|\mu\|_{TV} + \iota_{\operatorname{ran} L}(\mu)$*. Then, the following duality relationship holds:*

$$\min_{u \in \mathcal{B}} J(u) = \sup_{q \in \mathbb{R}^m, A^* q \in \operatorname{ran} L^*} -h^*((L^+)^* A^* q) - f_b^*(q). \tag{14}$$

In the special case $\operatorname{ran} L = \mathcal{M}$*, this yields*

$$\min_{u \in \mathcal{B}} J(u) = \sup_{q \in \mathbb{R}^m, A^* q \in \operatorname{ran} L^*, \|(L^+)^* A^* q\|_\infty \leq 1} -f_b^*(q). \tag{15}$$

In addition, let $(\hat{u}, \hat{q})$ *denote any primal-dual pair of problem* (14)*. The following duality relationships hold:*

$$A^* \hat{q} \in L^* \partial(\|\cdot\|_{TV})(L\hat{u}) \text{ and } -\hat{q} \in \partial f_b(A\hat{u}). \tag{16}$$

For a general operator L, computing h^* may be out of reach, since the conjugate of a sum cannot be easily deduced from the conjugates of each function in the sum. Hence, we now focus on problem (15) corresponding to the case $\operatorname{ran} L = \mathcal{M}$. This covers at least the two important cases $L = \mathrm{Id}$ and $L = D$, as shown in examples 2.3.1 and 2.3.4.

Remark 5. *In general, the dual problem does not need to have a solution. A straightforward application of [5, Th. 4.2] however shows that if either of the two following conditions hold*

1. $\operatorname{ran} A$ *intersects the relative interior of* $\operatorname{dom} f_b = \{q \in \mathbb{R}^m : f_b(q) < \infty\}$*,*
2. f_b *is polyhedral (i.e. has a convex polyhedral epigraph) and* $\operatorname{ran} A$ *intersects* $\operatorname{dom} f_b$*,*

the dual problem does have a solution. These conditions are mild: For all of the convex examples discussed in Section 2.2, the existence of a $u \in \mathcal{B}$ *with* $Au = b$ *is sufficient for at least one of them to hold.*

Solving the dual problem (15) does not directly provide a solution for the primal problem (1). The following proposition shows that it however yields information about the support of $L\hat{u}$, which is the critical information to retrieve.

Proposition 6. *Assume that* $\operatorname{ran} L = \mathcal{M}$ *and let* $(\hat{u}, \hat{q})$ *denote a primal-dual pair of problem* (15)*. Let* $I(\hat{q}) = \{x \in \Omega, |(L^+)^*(A^*\hat{q})|(x) = 1\}$*. Then*

$$\operatorname{supp}(L\hat{u}) \subseteq I(\hat{q}). \tag{17}$$

In particular, if $I(\hat{q}) = \{x_1, \ldots, x_p\}$*, then* $\hat{u}$ *can be written as:*

$$\hat{u} = u_K + \sum_{k=1}^{p} d_k L^+ \delta_{x_k} \tag{18}$$

with $u_K \in \ker L$ *and* $(d_k) \in \mathbb{R}^p$*. If problem* (1) *admits a unique solution, then* $p \leq \overline{m}$ *and* $\hat{u}$ *is the solution in Theorem 1.*

In the case where $I(\hat{q})$ is a finite set, Proposition 6 can be used to recover a solution $\hat{u}$ from $\hat{q}$, by injecting the specific structure (18) into (1). Let $(\lambda_i)_{1\leq i\leq r}$ denote a basis of $\ker L$ and define the matrix

$$M = \left[(\langle a_i, \lambda_k\rangle)_{1\leq i\leq m, 1\leq k\leq r}, (\langle (L^+)^* a_i, \delta_{x_j}\rangle)_{1\leq i\leq m, 1\leq j\leq p}\right] \tag{19}$$

Then problem (1) becomes a finite dimensional convex program which can be solved with off-the-shelf algorithms:

$$\min_{c\in\mathbb{R}^r, d\in\mathbb{R}^p} f_b\left(M\begin{bmatrix} c \\ d \end{bmatrix}\right) + \|d\|_1. \tag{20}$$

Overall, this section suggests the following strategy to recover $\hat{u}$:

1. Find a solution $\hat{q}$ of the finite dimensional dual problem (15).
2. Identify the support $I(\hat{q}) = \{x \in \Omega, |(L^+)^*(A^*\hat{q})|(x) = 1\}$.
3. If $I(\hat{q})$ is finitely supported, solve the finite dimensional primal problem (20) to construct $\hat{u}$.

Each step within this algorithmic framework however suffers from serious issues:

Problem 1 the dual problem (15) is finite dimensional but involves two infinite dimensional convex constraints sets

$$\mathcal{Q}_1 = \{q \in \mathbb{R}^m, A^*q \in \operatorname{ran} L^*\} \tag{21}$$

and

$$\mathcal{Q}_2 = \{q \in \mathbb{R}^m, \left\|(L^+)^* A^* q\right\|_\infty \leq 1\}, \tag{22}$$

which need to be handled with a computer.

Problem 2 finding $I(\hat{q})$ again consists of a possibly nontrivial maximization problem.

Problem 3 the set $I(\hat{q})$ may not be finitely supported.

To the best of our knowledge, finding general conditions on the functions

$$\rho_i = (L^+)^* a_i \tag{23}$$

allowing to overcome those hurdles is an open problem. It is however known that certain family of functions including polynomials and trigonometric polynomials [22] allow for a numerical resolution. In the following two sections, we study two specific cases useful for applications in details: the piecewise linear functions and trigonometric polynomials.

2.4.2 Piecewise linear functions in arbitrary dimensions

In this section, we assume that Ω is a bounded polyhedral subset of $\mathbb{R}^d$ and that each $\rho_i = (L^+)^* a_i$ is a piecewise linear function, with finitely many regions, all being polyhedral. This class of functions is commonly used in the finite element method. Its interest lies in the fact that any smooth function can be approximated with an arbitrary precision by using mesh refinements.

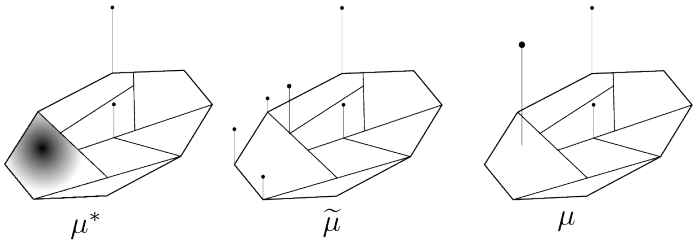

Figure 1: A graphical depiction of the three types of solutions for piecewise linear measurements.

Solving the primal For this class, notice that the function $(L^+)^* A^* q = \sum_{i=1}^m q_i \rho_i$ is still a piecewise linear function with finitely many polyhedral pieces. The maximum of the function has to be attained in at least one of the finitely many vertices $(v_j)_{j\in J}$ of the pieces. This is a key observation from a numerical viewpoint since it simultaneously allows to resolve problems 1 and 2. First, the constraint set $\mathcal{Q}_2$ can be described by a finite set of linear inequalities:

$$-1 \leq (L^+)^* A^* q(v_j) \leq 1, \quad j \in J.$$

Secondly, $I(\hat{q})$ can be retrieved by evaluating $(L^+)^* A^* q$ only on the vertices $(v_j)_{j\in J}$.

Unfortunately, problem 3 is particularly important for this class: $I(\hat{q})$ needs not be finitely supported since the maximum could be attained on a whole face. The following proposition however confirms that there always exists solutions supported on the vertices.

Proposition 7. *Suppose that* $\operatorname{ran} L = \mathcal{M}$ *and that the dual problem* (15) *has a solution. Then Problem* (1) *has at least one solution of the form*

$$\hat{u} = \sum_{j\in J} d_j L^+ \delta_{v_j} + u_K \tag{24}$$

with $u_K \in \ker L$, $d_j \in \mathbb{R}$ *and the* v_j *are the vertices of the polyhedral pieces.*

Once again, knowing the locations v_j of the Dirac masses in advance permits to solve (20) directly (without solving the dual) in order to obtain an exact solution of (1).

Sparsifying the solution For piecewise linear measurement functions, it turns out that the solution is not unique in general and that the form (24) is not necessarily the sparsest one. A related observation was already formulated in a different setting in [15], where the authors show that in 1D, two Dirac masses are usually found when only one should be detected. Figure 1 illustrates different types of possible solutions for a 2D mesh.

The proof of Proposition 7 suggests that one can sparsify a solution found by solving the primal problem resulting from the discretization through sampling on the grid of vertices. The basic reason is that piecewise linear measurements specify the zero-th and first order moments of a measure restricted to one piece. Among the infinitely many measures having these moments, one can pick the

sparsest one, consisting of a unique Dirac mass. This principle allows to pass from the 6-sparse measure $\tilde{\mu}$ to the 3-sparse measure μ in Figure 1.

To be precise, a collection of peaks $(d_i\delta_{x_i})_{i\in I}$, where

1. $\mathrm{conv}(x_i)_{i\in I}$ is contained in one polyhedral region of $\mathcal{G}$.
2. $(d_i)_{i\in I}$ have the same sign $\epsilon \in \{-1, 1\}$

can be combined into one peak $\overline{d}\delta_{\overline{x}}$, with

$$\overline{d} = \sum_{i\in I} d_i, \quad \overline{x} = \frac{1}{\overline{d}} \sum_{i\in I} d_i x_i.$$

We will see in the numerical experiments that this seemingly novel principle allows exact recovery of u under certain conditions on its initial structure.

Relationship to standard discretization The traditional way to discretize total variation problems with $L = \mathrm{Id}$ consists in imposing the locations of the Dirac masses on a set of predefined points $(x_i)_{1\leq i\leq n} \in \Omega^n$. Then, one can look for a solution of the form $\hat{u} = \sum_{i=1}^n d_k\delta_{x_i}$ and inject this structure in problem 1. By using this reasoning, there is no reason to find the exact solution of the original infinite dimensional problem. Proposition (7) sheds a new light on this strategy, by telling that this actually amounts to solving *exactly* an infinite dimensional problem with piecewise linear measurement functions.

2.4.3 Trigonometric polynomials in 1D

In this section, we assume that $\Omega = \mathbb{T}$ is the one dimensional torus (see remark (1)). For $j \in \mathbb{N}$, let $p_j(t) = \exp(-2\imath\pi jt)$. We also assume that the functions ρ_i are real trigonometric polynomials:

$$\rho_i = \sum_{j=-K}^{K} \gamma_{j,i} p_j,$$

with $\gamma_{j,i} = \gamma^*_{-j,i}$. For this problem, the strategy suggested in section 2.4.1 will be adopted.

Solving the dual The following simple lemma states that in the case of a finite dimensional kernel, the constraint set $\mathcal{Q}_1$ is just a finite dimensional linear constraint.

Lemma 3. *Let $r = \dim(\ker(L)) < \infty$ and $(\lambda_i)_{1\leq i\leq r}$ denote a basis of $\ker L$. The set $\mathcal{Q}_1$ can be rewritten as*

$$\mathcal{Q}_1 = \{q \in \mathbb{R}^m, \forall 1 \leq i \leq r, \langle q, A\lambda_i\rangle = 0\}.$$

Proof. Since $\mathrm{ran}\, L^* = V$ (by the closed range theorem), $A^*q \in \mathrm{ran}\, L^*$ if and only if $\forall 1 \leq i \leq r, \langle A^*q, \lambda_i\rangle = 0$. □

Hence, when $\ker L$ is finite-dimensional the set $\mathcal{Q}_1$ can be easily handled by using numerical integration procedures to compute the mr scalars $\langle a_k, \lambda_i\rangle$. Let us now turn to the set $\mathcal{Q}_2$. The following lemma is a simple variation of [14, Thm 4.24]. It was used already for super-resolution purposes [8].

Lemma 4. *The set $\mathcal{Q}_2$ can be rewritten as follows:*

$$\mathcal{Q}_2 = \left\{ \alpha \in \mathbb{R}^m, \exists Q \in \mathbb{C}^{(2K+1)\times(2K+1)}, \begin{bmatrix} Q & \Gamma\alpha \\ (\Gamma\alpha)^* & 1 \end{bmatrix} \succeq 0, \right.$$
$$\left. \sum_{i=1}^{2K+2-j} Q_{i,i+j} = \begin{cases} 1, & j = 0, \\ 0, & 1 \leq j \leq 2K+1. \end{cases} \right\}.$$

With Lemmas 3 and 4 at hand, the dual problem (15) becomes a semidefinite program that can be solved with a variety of approaches, such as interior point methods [31].

Finding the Dirac mass locations The case of trigonometric polynomials makes Proposition 6 particularly helpful. In that case, either the trigonometric polynomial is zero and the solution $\hat{u}$ lives in the kernel of L, or the set I is finite with cardinality at most $2K$, since $|(L^+)^* A^* q|^2 - 1$ is a negative trigonometric polynomial of degree $4K + 2$. Retrieving its roots can be expressed as an eigenvalue evaluation problem [11] and be solved efficiently.

3 Numerical Experiments

In this section, we perform a few numerical experiments to illustrate the theory. In all our experiments, we use the toolbox CVX [23] for solving the resulting convex minimization problems.

3.1 Piecewise linear functions

3.1.1 Identity in 1D

In this paragraph, we set $L = \mathrm{Id}$ and $\Omega = [0, 1]$. We assume that the functions a_i are random piecewise linear functions on a regular grid. The values of the functions on the vertices are taken as independent random Gaussian realizations with standard deviation 1. In this experiment, we set u as a sparse measure supported on 3 points. We probe it using 12 random measurement functions a_i and do not perturb the resulting measurement vector b, allowing to set $f_b = \iota_{\{b\}}$. The result is shown on Figure 2. As can be seen, the initially recovered measure is 7 sparse. Using the sparsification procedure detailed in paragraph 2.4.2 allows to *exactly* recover the true 4 sparse measure u. We will provide a detailed analysis of this phenomenon in a forthcoming paper.

3.1.2 Derivative in 1D

In this section we set $\Omega = [0, 1]$ and $L = D$. We assume that the functions a_i are piecewise constant. In the terminology of [30], this means that we are *sampling splines with splines*. By equation (13), we see that the functions $\rho_i = (L^+)^* a_i$ are piecewise linear and satisfy $\rho_i(0) = \rho_i(1) = 0$.

In this example, we set the values of a_i on each piece as the realization of independent normally distributed random variables. We divide the interval $[0, 1]$ in 10 intervals of identical length. An example of a sampling function is displayed in Figure 3.

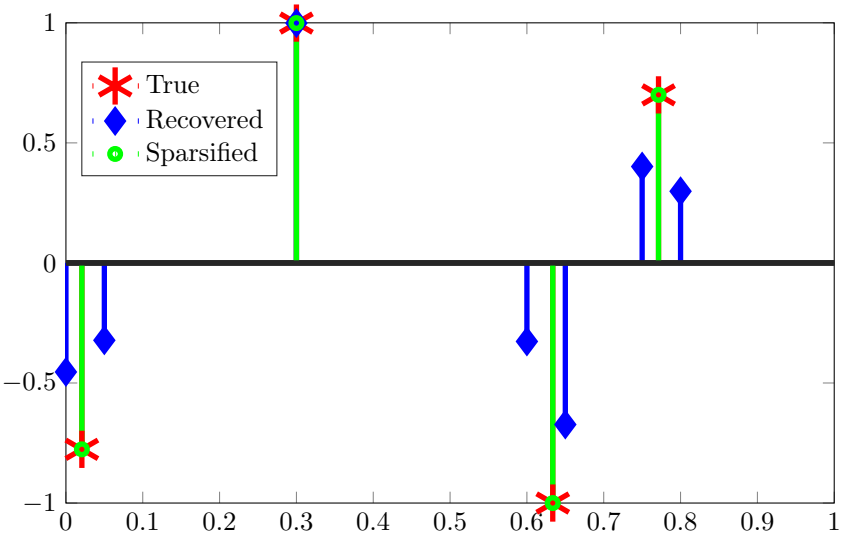

Figure 2: Example of recovery with random piecewise linear measurement functions in 1D. The solution recovered by a standard ℓ^1 solver is not the sparsest one. The sparsification procedure proposed in the paper allows recovering the sparsest solution and recovering *exactly* the sampled function.

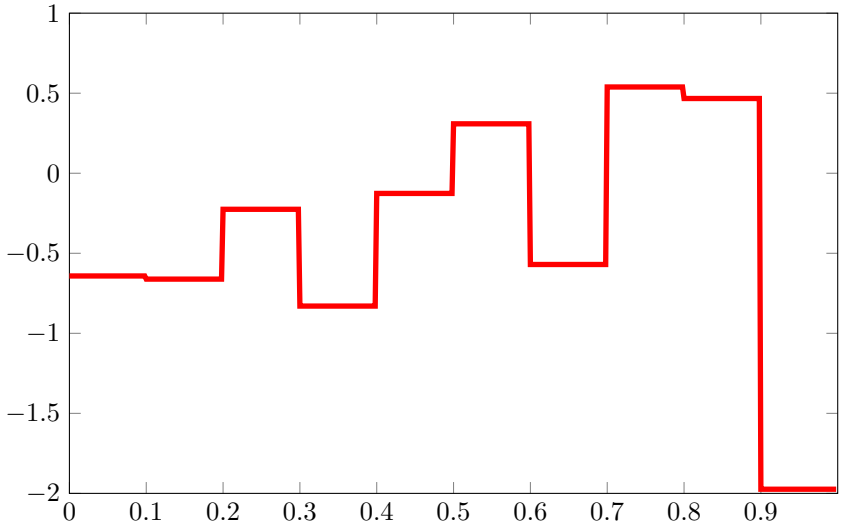

Figure 3: Sampling function a_1 used to probe piecewise constant signals. The others have a similar structure with other random values on each interval.

The sensed signal u is defined as piecewise constant with jumps occurring outside the grid points. Its values are comprised in $[-1, 1]$.

The measurements are obtained according to the following model: $b_i = \langle a_i, u \rangle + \epsilon_i$, where ϵ_i is the realization of a Bernoulli-Gaussian variable. It takes the value 0 with probability 0.9 and takes a random Gaussian value with variance 3 with probability 0.1. To cope with the fact that the noise is impulsive, we propose to solve the following problem ℓ^1 fitted and total variation regularized problem.

$$\min_{u \in BV(]0,1[)} \|Du\|_{TV} + \alpha \|Au - b\|_1, \tag{25}$$

where $\alpha = 1$.

A typical result of the proposed algorithms is shown in Figure 4. Here, we probe a piecewise constant signal with 3 jumps (there is a small one in the central plateau) with 42 measurements. Once again, we observe perfect recovery despite the additive noise. This favorable behavior can be explained by the fact that the noise is impulsive and by the choice of an ℓ^1 data fitting term.

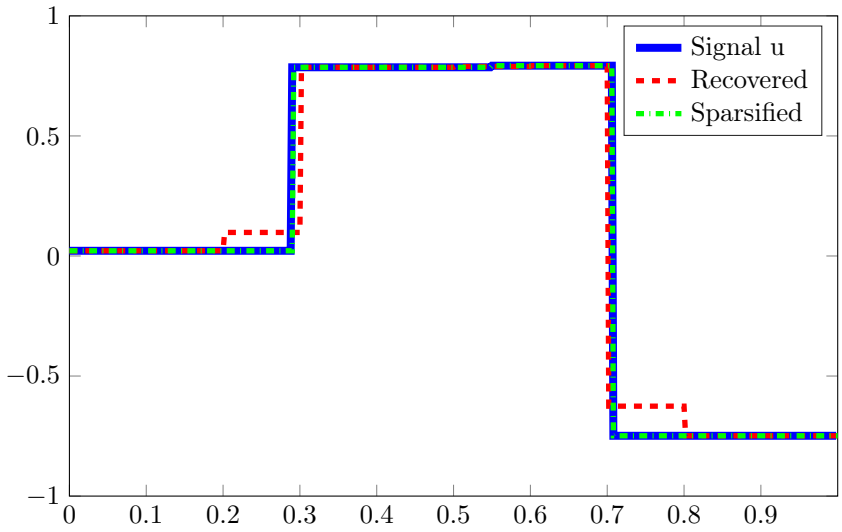

Figure 4: Example of recovery of a piecewise linear signal u with measurements corrupted by Bernoulli-Gaussian noise. Once again, the proposed algorithm implemented with the proposed sparsification procedure recovers the true signal exactly, despite noise.

3.1.3 Identity in 2D

In this section, we set $\Omega = [0, 1]^2$ and $L = \mathrm{Id}$. We probe a sparse measure $\mu \in \mathcal{M}([0, 1]^2)$ using real trigonometric polynomials up to order 5. We then solve the problem 1 with f_b modeling box-constraints and A being the measurement operator associated with the piecewise functions formed by linearizing the trigonometric polynomials on a regular grid $\{0, 0.1, \ldots, 1\}^2$. Then, we collapse the resulting solution into a sparser one. To avoid numerical problems, we discarded all peaks with an amplitude less than 10^{-8} before the last step. The

results, together with an illustration of the collapsing procedure, are depicted in Figure 3.1.3.

3.2 Trigonometric polynomials.

We generate $m = 35$ trigonometric polynomials

$$a_i(t) = \sum_{j=-N}^{N} \gamma_{j,i} \exp(-2\iota\pi jt)$$

of degree $N = 50$ as follows: for $j \geq 0$, we set the coefficients $\gamma_{j,i}$ of the i:th polynomial to be

$$\gamma_{j,i} = \frac{\xi_{j,i}}{\max(j,1)},$$

where $\xi_{j,i}$ are i.i.d. normal distributed. For $j < 0$, we set $\gamma_{j,i} = \gamma^*_{-j,i}$. This ensures that the functions a_i are real, and furthermore have a good approximation rate with respect to trigonometric polynomials. Seven such functions are depicted in Figure6. Note that we do not need to worry about a_i not vanishing at $\pm 1/2$, since the functions live on the torus, a manifold without boundary.

We then generate $b \in \mathbb{R}^m$ by measuring a ground truth measure $\mu_0 = \sum_{i=-2}^{2} c_i \delta_{x_i}$, where x_i are chosen as

$$x_i = \frac{i}{5} + n_i,$$

where n_i are small random displacements, and $i.i.d$ normally distributed amplitudes $(c_i)_{i=1}^5$. Next, for each $K = 10, 11, \ldots, 50$, we solve the problem 1, with A being the measurement operator with respect to the functions

$$\widetilde{a}_i^K(t) = \sum_{j=-K}^{K} \gamma_{j,i} \exp(-2\iota\pi jt).$$

In Figure 7, we plot the results of the minimization (1) with

$$f_b(x) = \frac{100 \left\|x\right\|_2^2}{2},$$

depending on K. We see that already for $K = 30$, the solution is reasonably close to the true solution (at $N = 50$) (the relative error in the input, $\left\|\widetilde{A}\mu_0 - b\right\|_2 / \left\|b\right\|_2$, for this K is approximately equal to 0.06). The latter is furthermore essentially equal to the ground truth μ_0.

4 Proofs

In this section, we include all proofs left out in the main text.

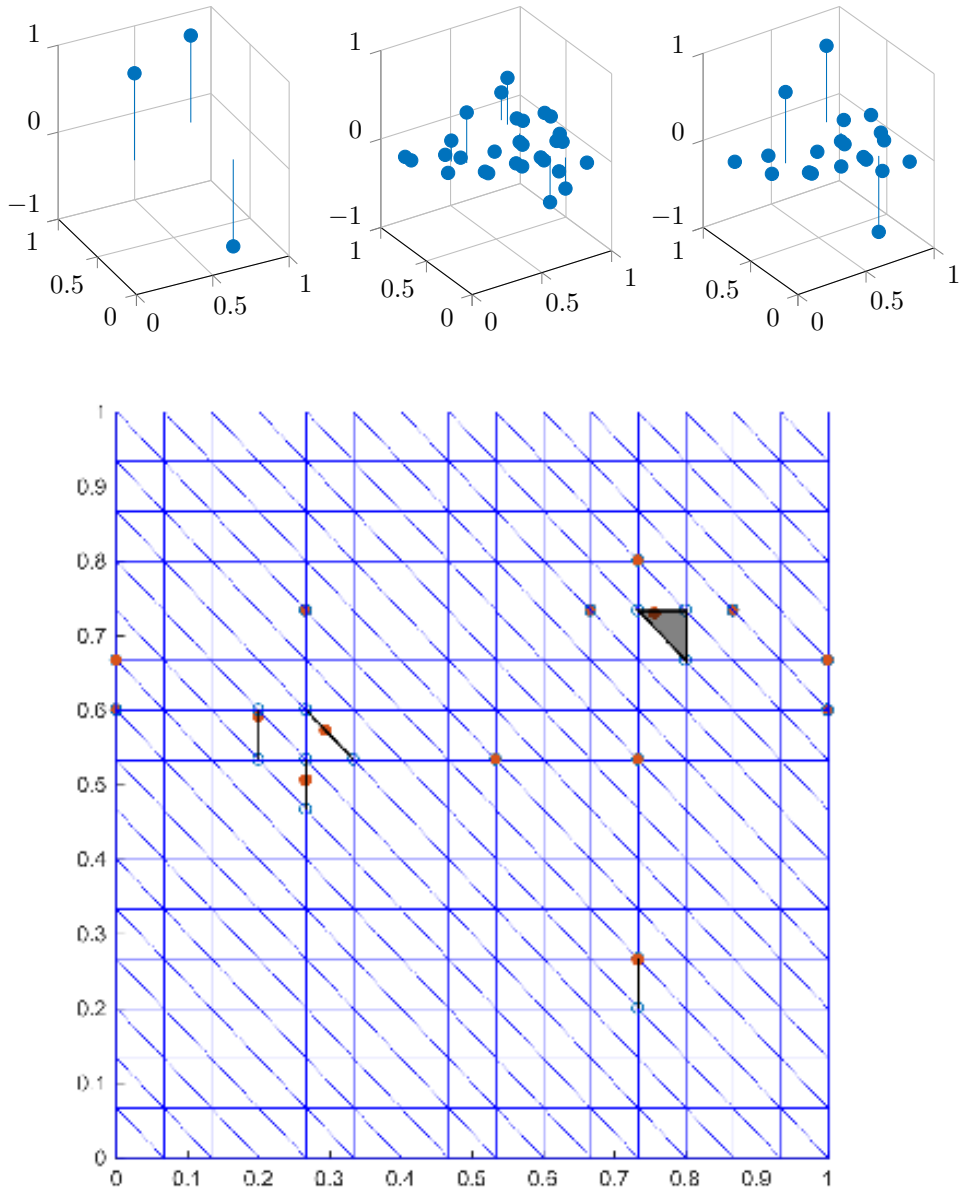

Figure 5: Top: The ground truth measure, the solution obtained by sampling on the vertices, and the sparsified solution. Bottom: Illustration of the sparsification procedure. The circles represent the initial solution, while the dots indicate the sparsified solution. A thick trait or a grey region indicates masses that have been merged.

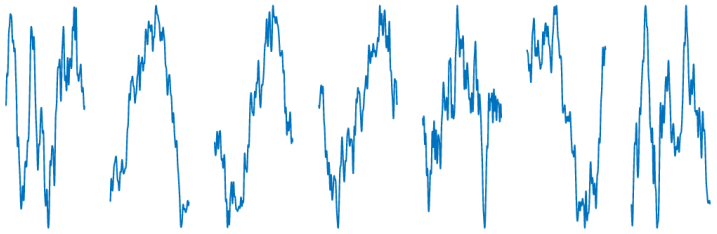

Figure 6: Seven randomly generated trigonometric polynomials a_i.

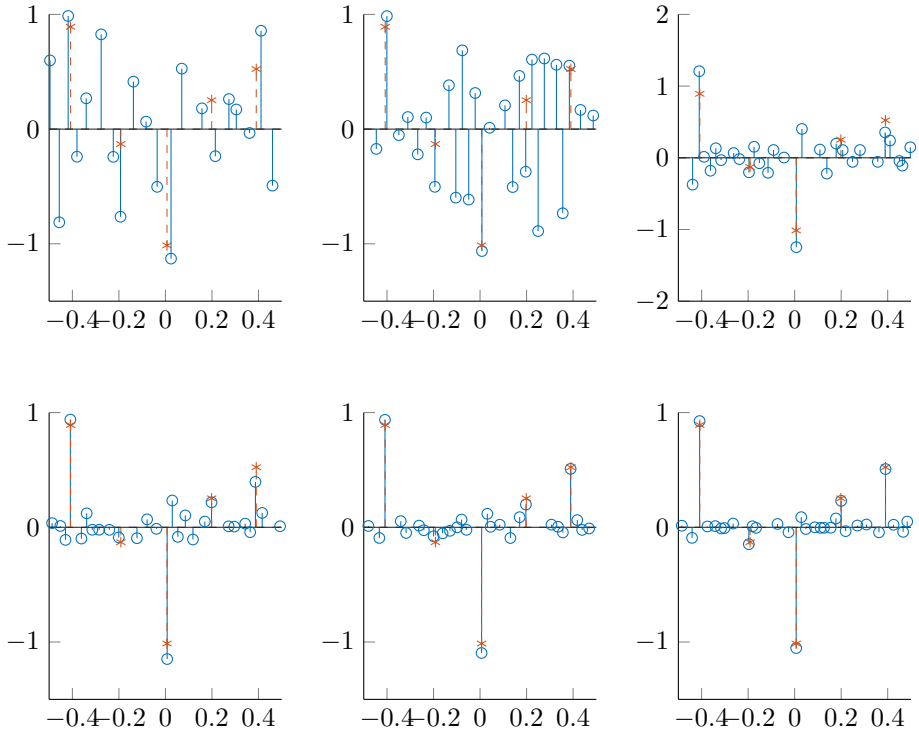

Figure 7: Minimizers of (1) ($*$) together with the ground truth μ_0 ($\circ$) for (from above left to below right) K equal to $15, 20, 25, 30, 35$ and 40, respectively.

4.1 Structure of solutions

As was argued already in the main body of the text, the proof of Theorem 1 can be broken down to a treatment of the problem (4). In the following, we will carry out the argument proving that the latter problem has a solution of the claimed form.

We first prove the result in finite dimensions and then use a limit argument. The statement is well known in finite dimension, see e.g. [29, Theorem 6] and [27]. We provide a proof for completeness. It has a geometrical flavour.

Lemma 5. *Let $m, n \in \mathbb{N}$, $G \in \mathbb{R}^{m,n}$, $b \in \operatorname{ran} G$ and $m \leq n$. Then a problem of the form*

$$\min_{u \in \mathbb{R}^n} \|u\|_1 \quad \textit{subject to } Gu = b. \tag{26}$$

has a solution $\hat{u}$ of (1) of the form

$$\hat{u} = \sum_{k=1}^{p} c_k e_{i_k},$$

with $(c_k)_{k=1}^m$ some real scalars and $p \leq m$.

Proof. Let u be a solution to (26) (its existence easily follows from the coercivity of the 1-norm and the non-emptiness and closedness of the set $G^{-1}(\{b\})$). The image $b = Gu$ then lies on the boundary of the polytope $P = G\{u \mid \|u\|_1 \leq \|u\|_1\}$ – if it did not, b would be of the form $G\tilde{u}$ with $\|\tilde{u}\|_1 < \|u\|_1$. Then $\tilde{u}$ would be a feasible point with smaller objective value than u, which is a contradiction to the optimality of u.

The polytope P is at most m-dimensional, hence its boundary ∂P consists of faces of dimension at most $m-1$. Having just argued that b lies on that boundary, it must lie on one of those faces, say F, which then has dimension at most $m-1$. Concretely, $b \in \operatorname{conv}(\operatorname{vert}(F))$, where $\operatorname{vert}(F)$ denotes the set of vertices of face F. The vertices of F are the images by G of a subset of the ℓ_1-ball's vertices, so they can be written as $\|u\|_1 \epsilon_k G e_i$, for some $i \in \{1, \dots, n\}$ and for $\epsilon_k \in \{-1, 1\}$. Caratheodory's theorem applied in the $(m-1)$-dimensional space $\operatorname{aff} F$ implies that b can be written as

$$b = \sum_{k=1}^{m} \theta_k \|\hat{u}\|_1 \epsilon_k G e_{i_k}$$

with $\sum_{k=1}^m \theta_k = 1$ and $\epsilon \in \{\pm 1\}^m$. The vector $\|u\|_1 \sum_{k=1}^m \theta_k \epsilon_k e_{i_k}$ is a solution of (26) of the stated form. □

The strategy will now be to discretize the problem on finer and finer grids, use the previous lemma and pass to the limit.

Lemma 6. *Define a sequence of discretizations $(\Omega_n)_{n \in \mathbb{N}}$ of Ω as*

$$\Omega_n = \left([-2^n, 2^n]^d \cap \frac{\mathbb{Z}^d}{2^n} \right) \cap \Omega. \tag{27}$$

For $k \in \Omega_n$, define ω_n^k to be the hypercube of center k and side-length 2^{-n} intersected with Ω. Let $\mu \in \mathcal{M}(\Omega)$ denote a measure and define the sequence:

$$\nu_n = \sum_{k \in \Omega_n} \mu(\omega_k)\delta_k. \tag{28}$$

Then $\nu_n \overset{*}{\rightharpoonup} \mu$ and $\|\nu_n\|_{TV} \leq \|\mu\|_{TV}$.

Proof. First, it follows directly from the definition of the total variation that

$$\|\nu_n\|_{TV} = \sum_{k \in \Omega_n} |\mu(\omega_k)| \leq \|\mu\|_{TV}. \tag{29}$$

We now need to prove that for each $\phi \in \mathcal{M}^*$, $\langle \nu_n, \phi \rangle \to \langle \mu, \phi \rangle$. So fix ϕ and let $\epsilon > 0$. Since $\phi \in \mathcal{M}^*$, there exists a compact set K with the property $|\phi(x)| < \epsilon$ for $x \notin K$. Since ϕ is equicontinuous on K, there exists a $\delta > 0$ so that if $\|x - x'\|_\infty < \delta$, $|\phi(x) - \phi(x')| < \epsilon$. If we choose n so large so that $2^{-n} < \delta$, we will have

$$\begin{aligned}
|\langle \mu - \nu_n, \phi \rangle| &\leq \int_{\Omega \setminus K} |\phi| \, d(|\mu| + |\nu_n|) + \left| \int_K \phi d\mu - \int_K \phi d\nu_n \right| \\
&\leq \epsilon(\|\mu\|_{TV} + \|\nu_n\|_{TV}) + \left| \sum_{k \in \Omega_n} \int_{\omega_k} \phi d\mu - \phi(k)\mu(\omega_k) \right| \\
&\leq 2\epsilon \|\mu\|_{TV} + \sum_{k \in \Omega_n} \int_{\omega_k} |\phi(\ell) - \phi(k)| \, d\mu(\ell) \\
&\leq 2\epsilon \|\mu\|_{TV} + \epsilon \sum_{k \in \Omega_n} |\mu(\omega_k)| \\
&\leq 3\epsilon \|\mu\|_{TV}.
\end{aligned}$$

Since $\epsilon > 0$ was arbitrary, the claim follows. □

When passing to the limit in our limit argument, we we will need the following continuity property of the operator AL^+:

Lemma 7. *The operator $AL^+ : \mathcal{M} \to \mathbb{R}^m$ is weak-$*$-weak continuous. That is, if $\mu_n \overset{*}{\rightharpoonup} \mu$, $AL^+\mu_n \to AL^+\mu$. The same is true for $H = \Pi_{X^\perp} A$.*

Proof. We simply need to note that $\mu_n \overset{*}{\rightharpoonup} \hat{\mu}$ and assumption 3 implies that

$$\left\langle a_i, L^+\mu_n \right\rangle = \left\langle (L^+)^* a_i, \mu_n \right\rangle = \langle \mu_n, \rho_i \rangle \to \langle \mu, \rho_i \rangle = \left\langle a_i, L^+\mu \right\rangle = \left\langle (L^+)^* a_i, \mu \right\rangle, \tag{30}$$

and that $\Pi_{X^\perp}$ is continuous. □

Now let us prove that the optimal value of the problem (1) can be found by solving slightly perturbed discretized problems.

Lemma 8. *Let $b \in \operatorname{ran} H$. There exists a sequence $(b_n)_{n \in \mathbb{N}}$ of vectors in $\mathbb{R}^m$ with the following properties*

- *For each n, b_n is in the range of the n-th discretized H-operator, i.e.*

$$b_n \in H \operatorname{span} (\delta_\omega)_{\omega \in \Omega_n}$$

- b_n *converges to* b.
- *For* $n \in \mathbb{N}$, *define* $\hat{J}_n$ *through*

$$\hat{J}_n := \min_{c \in \mathbb{R}^{|\Omega_n|}} \|c\|_1 \text{ subject to } HL^+ \left(\sum_{k \in \Omega_n} c_k \delta_k \right) = b_n. \qquad (\mathcal{P}_n)$$

Then $\liminf_{n\to\infty} \hat{J}_n \leq \hat{J}$, *where* $\hat{J}$ *is the optimal value of problem* (4).

Proof. First, we note that problem (4) has a solution $\hat{\mu}$. We skip the proof since it is identical to that of Proposition 1.

Now, according to Lemma 6, there exists a sequence of measures μ_n of the form

$$\mu_n = \sum_{k \in \Omega_n} c_k \delta_k$$

with $\mu_n \overset{*}{\rightharpoonup} \hat{\mu}$ and $\|\mu_n\|_{TV} \leq \|\hat{\mu}\|_{TV}$ for each n.

Lemma 7, again together with the continuity of $\Pi_{X^\perp}$, now implies that $H\mu_n \to H\hat{\mu} = b$. If we put $b_n = AL^+\mu_n$, b_n is in the range of the n-th discretized A-operator, $b_n \to b$, and $\|\mu_n\|_{TV} \geq \hat{J}_n$. This implies

$$\liminf_{n\to\infty} \hat{J}_n = \liminf_{n\to\infty} \|\mu_n\|_{TV} \leq \|\hat{\mu}\|_{TV} = \hat{J}.$$

□

We may now prove the main result of this section.

Proof of Theorem 1. By definition $b \in \operatorname{ran} H$. We can hence apply Lemma 8 to construct a sequence $(b_n)_{n\in\mathbb{N}}$ having the properties stated in the mentioned Lemma.

Now consider the problems $(\mathcal{P}_n)$. If we write them down explicitely, we see that the minimization over the vectors $c^{(n)}$ are exactly as in Lemma 5, with $G = H$ and $m = \overline{m}$. Hence, we can construct a sequence $(\hat{c}_n)$ of solutions, where $\hat{c}_n$ containing $p_n \leq \overline{m}$ nonzero components for $n \geq \overline{m}$. Thus, we may write

$$\sum_{k \in \Omega_n} \hat{c}_{n,k} \delta_k = \sum_{\ell=1}^{\overline{m}} d_{n,\ell} \delta_{x_{n,\ell}},$$

for some $d_n \in \mathbb{R}^{\overline{m}}$ and $X_n = (x_{n,l})_l \in \Omega^{\overline{m}}$. In case $p_n < \overline{m}$, we may repeat positions in the vector X_n.

Now $(d_n)_{n\in\mathbb{N}}$ is bounded, since $\|d_n\|_1 \leq \hat{J}_n \leq \hat{J}_1$ for each n. This implies that there exists a subsequence, which we do not rename, such that d_n is converging to $d^* \in \mathbb{R}^{\overline{m}}$. By possibly considering a subsequence of this subsequence, we may assume that X_n converges in $\overline{\Omega}^\times$, where $\overline{\Omega}^\times$ denotes the one-point-compactification Ω. This means that each of the component sequences $(x_{n,\ell})_n$ either converges to a point x^*_ℓ in Ω, or diverges to ∞, meaning that it escapes every compact subset of Ω.

Consequently, the subsequence $\mu_n = \sum_{\ell=1}^{\overline{m}} d_{n,\ell} \delta_{x_{n,\ell}} \overset{*}{\rightharpoonup} \sum_{\ell=1}^{\overline{m}} d^*_\ell \delta_{x^*_\ell} =: \mu^*$, where we identify δ_∞ with the zero measure (note that if $x_{n,\ell} \to \infty$, then $\delta_{x_{n,\ell}} \overset{*}{\rightharpoonup} 0$)).

Lower semi-continuity of the TV-norm implies

$$\left\| \sum_{\ell=1}^{\overline{m}} d_\ell^* \delta_{x_\ell^*} \right\|_{TV} \leq \liminf_{n\to\infty} \left\| \sum_{\ell=1}^{p} d_\ell \delta_{x_{n,\ell}} \right\|_{TV} = \liminf_{n\to\infty} \hat{J}_n \leq \hat{J},$$

where we used Lemma 6 in the final step. Also, applying Lemma 7 together with the properties of (b_n), we get

$$H\mu^* = \lim_{n\to\infty} H\mu_n = \lim_{n\to\infty} b_n = b.$$

Hence, $\sum_{\ell=1}^{\overline{m}} d_\ell^* \delta_{x_\ell^*}$ is a solution of (1), which was exactly what was needed to be proven. (Note that any $x_\ell^* = \infty$ will only cause the linear combination of δ-peaks to be shorter). □

4.2 Numerical Resolution

In this section, we prove the propositions stated in Section 2.4. We begin with the one describing the dual problem of (1).

Proof of Proposition 5. Define $g : \mathcal{B} \to \mathcal{M}$ with $g(u) := \|Lu\|_{TV}$. Then $J(u) = f_b(Au) + g(u)$. Standard duality arguments [16, p.60] yield:

$$\min_{u\in\mathcal{B}} J(u) = \sup_{q\in\mathbb{R}^m} -g^*(-A^*q) - f_b^*(q). \tag{31}$$

Now, we have:

$$\begin{aligned}
g^*(z) &= \sup_{u\in\mathcal{B}} \langle z, u\rangle - g(u) \\
&= \sup_{u\in\mathcal{B}} \langle z, u\rangle - \|Lu\|_{TV} \\
&= \sup_{v\in V, u_K\in\ker L} \langle z, v + u_K\rangle - \|Lv\|_{TV} \\
&= \begin{cases} \sup_{v\in V} \langle z, v\rangle - \|Lv\|_{TV} & \text{if } z \in (\ker L)^\perp, \\ +\infty & \text{otherwise} \end{cases} \\
&= \begin{cases} \sup_{v\in V} \langle z, L^+Lv\rangle - \|Lv\|_{TV} & \text{if } z \in \operatorname{ran} L^*, \\ +\infty & \text{otherwise} \end{cases} \\
&= \begin{cases} \sup_{w\in\operatorname{ran} L} \langle (L^+)^* z, w\rangle - \|w\|_{TV} & \text{if } z \in \operatorname{ran} L^*, \\ +\infty & \text{otherwise} \end{cases} \\
&= \begin{cases} h^*((L^+)^* z) & \text{if } z \in \operatorname{ran} L^*, \\ +\infty & \text{otherwise} \end{cases}
\end{aligned}$$

We used the closed range theorem, which in particular implies that $\operatorname{ran} L^* = (\ker L)^\perp$ for an operator L with closed range.

For the special case of $\operatorname{ran} L = \mathcal{M}$, we note that

$$h^*(\phi) = \sup_{\mu\in\mathcal{M}} \langle \phi, \mu\rangle - \|\mu\|_{TV} = \begin{cases} 0 & \text{if } \|\phi\|_\infty \leq 1. \\ +\infty & \text{otherwise,} \end{cases}$$

Note that the subdifferential of g at every $u \in \mathcal{B}$ reads $\partial g(u) = L^* \partial(\|\cdot\|_{TV})(Lu)$ (see e.g. [16, Prop.5.7]. The duality relations also follows from standard arguments, see e.g. [16, p.60]. □

Next, we prove the proposition describing how to construct a primal solution from a dual one in the case that $\operatorname{ran} L = \mathcal{M}$.

Proof of Proposition 6. We have for any operator L obeying assumption 2

$$(L^+)^* L^* = (LL^+)^* = \Pi^*_{\operatorname{ran} L} = \mathrm{Id},\ L^*(L^+)^* = (L^+L)^* = j_V^*.$$

By construction, $A^*\hat{q}$ and $L^*\partial(\|\cdot\|_{TV})(L\hat{u})$ are elements of $\operatorname{ran} L^*$. Due to the closed range theorem, $\operatorname{ran} L^*$ is isomorphic to the annihilator V. On that space, j_V^* is injective. Hence, the inclusion (16) is equivalent to

$$(L^+)^*(A^*\hat{q}) \in \partial(\|\cdot\|_{TV})(L\hat{u}). \tag{32}$$

Now, it is well known (see for instance [15]), that for all $\mu \in \mathcal{M}$,

$$\partial(\|\cdot\|_{TV})(\mu) = \left\{ \eta \in \mathcal{M}^*, \|\eta\|_\infty \leq 1, \int_\Omega \eta(t)\, d\mu(t) = \|\mu\|_{TV} \right\}. \tag{33}$$

Consequently, (32) tells us that the continuous function $(L^+)^*(A^*\hat{q})$ has modulus 1 $L\hat{u}$-almost everywhere on $\operatorname{supp}(L\hat{u})$. This means that

$$(L\hat{u})(I(\hat{q}) \backslash \operatorname{supp}(L\hat{u})) = 0.$$

In particular, if the set I only consists of isolated points, we get $\operatorname{supp}(L\hat{u}) \subseteq I$. Hence, there exists $(d_k)_{1\leq k\leq p}$ with

$$L\hat{u} = \sum_{k=1}^p d_k \delta_{x_k} \implies \hat{u} = u_K + \sum_{k=1}^p d_k L^+ \delta_{x_k} \tag{34}$$

for some $u_K \in \ker L$. □

Next, we prove the claim about the structure and possible numerical resolution of the optimal measure $\hat{\mu}$ in the case of piecewise linear ρ_i

Proof of Proposition 7. Let $\hat{q}$, and $L^+\mu^* + u_K$ be any solution of $\min_{u\in\mathcal{B}} J(u)$ with supp $\mu^* \subseteq \cup_{\ell=1}^n F_\ell$, where F_ℓ are the faces described above (such a solution exists due to Theorem 1). Standard duality arguments (see for instance [16, prop. 4.1]) yield that $\hat{q}$ and $L^+\mu^* + u_K$ satisfies the primal-dual conditions 16, i.e. in particular 32, since $\operatorname{ran} L = \mathcal{M}$.

It is clear that any atomic measure $\mu = \sum_{j=1}^n d_j \delta_{\overline{x}_j}$ with $\langle \rho_i, \mu^* \rangle = \langle \rho_i, \mu \rangle$ for each i and $\|\mu\|_{TV} = \|\mu^*\|_{TV}$, $L^+\mu^* + u_K$ also is a solution to $\min J(u)$. Such a measure can be constructed as follows: Suppose that there exists a face P of at least dimension 1 of a polytope F_j such that supp μ_* intersects in at least one point p of the relative interior of P (if no such P exists, μ_* is already atomic). Due to supp$\mu_* \subseteq I(\hat{q})$, $(L^+)^*A^*q$ has absolute value 1 in p. $(L^+)^*A^*q$ being a continuous piecewise linear function with absolute value bounded by one, it must therefore have a constant value ε, either equal to $+1$ or -1, on P. Due to the structure (33) of the subdifferential of the TV-norm, this implies that μ_*

(the unimodular part of the polar decomposition of μ_* to be exact) must have the same sign as ϵ almost everywhere on P.

On P, each function ρ_i can be written as $\rho_i(x) = \langle \alpha_{ij}, x\rangle + \beta_{ij}$, for some vectors $\alpha_{ij} \in \mathbb{R}^d$ and scalars $\beta_{ij} \in \mathbb{R}$. If we hence define

$$d = \epsilon \left|\mu\right|^*(P), \quad \overline{x} = \frac{1}{\left|\mu\right|^*(P)} \int_P x d\left|\mu\right|^* \in P,$$

and $\mu = d\overline{x} + \mu^*|_{\Omega\setminus P}$, we have $\|\mu\|_{TV} = |d| + \left\|\mu^*|_{\Omega\setminus P}\right\| = \|\mu^*|_P\|_{TV} + \left\|\mu|_{\Omega\setminus P}\right\| = \|\mu^*\|_{TV}$, and

$$\langle \rho_i, \mu\rangle = \sum_{j=1}^n \left(\langle \alpha_{ij}, \overline{x}_j\rangle + \beta_{ij}\right)\mu^*(F_j) = \sum_{j=1}^n \int_{F_j} \left(\langle \alpha_{ij}, x\rangle + \beta_{ij}\right) d\mu^* = \langle \rho_i, \mu^*\rangle.$$

By iteratively removing all such non-atomic parts of μ_*, we obtain an atomic solution μ.

We still need to prove that we can find a $\mu^* = \widetilde{\mu}$ which is atomic and supported on the vertices of F_j. Note that each $\overline{x}_j$ can be represented as a convex combination $\sum_{k=1}^{t_j} \theta_k v_{jk}$ of the vertices v_{jk} of F_j. Defining a measure

$$\widetilde{\mu} = \sum_{j=1}^n \sum_{k=1}^{t_j} \theta_k \mu(F_j) \delta_{v_{jk}},$$

we see that $\|\mu\|_{TV} = \|\widetilde{\mu}\|_{TV}$ and

$$\langle \rho_i, \widetilde{\mu}\rangle = \sum_{j=1}^n \sum_{k=1}^{t_j} \theta_k \left(\langle \alpha_{ij}, v_{jk}\rangle + \beta_{ij}\right)\mu(F_j) = \sum_{j=1}^n \left(\langle \alpha_{ij}, \overline{x}_j\rangle + \beta_{ij}\right)\mu(F_j) = \langle \rho_i, \mu\rangle,$$

so that $\widetilde{\mu}$ is also a solution.

□

Finally, we provide the argument that the constraint of the dual problem can be rewritten as an inequality on the space of Hermitian matrices in the case of the functions ρ_i begin trigonometric polynomials.

Proof of Lemma 4. Note that $|\sum_{i=1}^m \alpha_i \rho_i| \le 1$ is equivalent to

$$1 \ge \left|\sum_{i=1}^m \alpha_i \sum_{j=-K}^K \gamma_{i,j} p_j\right| = \left|\sum_{j=-K}^K \sum_{i=1}^m \alpha_i \gamma_{i,j} p_j\right| = \left|\sum_{j=-K}^K (\Gamma\alpha)_j p_j\right| = \left|p_{-K} \sum_{j=-K}^K (\Gamma\alpha)_j p_j\right|.$$

The function $f = p_{-K}\sum_{j=-K}^K (\Gamma\alpha)_j p_j$ is a causal trigonometric polynomial. We know from [14, Cor.4.27] that it obeys the constraint $\|f\|_\infty \le 1$ if and only if there exists a positive semi-definite matrix $Q \in \mathbb{C}^{(2K+1)\times(2K+1)}$ such that

$$\begin{bmatrix} Q & \Gamma\alpha \\ (\Gamma\alpha)^* & 1 \end{bmatrix} \succeq 0 \text{ and } \sum_{i=1}^{2K+2-j} Q_{i,i+j} = \begin{cases} 1, & j = 1, \\ 0, & 2 \le j \le 2K+1. \end{cases} \tag{35}$$

□

4.3 Differential operators of Section 2.3.3

Here, we provide the proofs for the lemmas and propositions which include more general differential operators into our framework. We begin by proving a preparatory lemma about the operator L^+ in (7).

Lemma 9. *The operator L^+ defined by (7) is a continuous operator from $\mathcal{M}(\Omega)$ to $\mathcal{C}(\Omega)$. It has the property $P(D)L^+ = \mathrm{Id}_{\mathcal{M}(\Omega)}$.*

Proof. Let us begin by showing that L^+ maps from $\mathcal{M}(\Omega)$ to $\mathcal{C}(\Omega)$. First, note that the continuity of $x \mapsto u_x$ implies that $L^+\mu$ is pointwise well-defined. We still need to show that for a fixed μ, the map $x \mapsto (L^+\mu)(x)$ is continuous. This follows from a standard "limits and integrals commute" argument. Let $x_n \to x$. Then $u_y(x_n) \to u_y(x)$ pointwise. Furthermore, $|u_y(x_n)| \leq \sup_{y\in\Omega} \|u_y\|_\infty$ for all y and x_n. Since $\sup_{y\in\Omega} \|u_y\|_\infty$ is a μ-integrable function, the theorem of Lebesgue implies that

$$\lim_{n\to\infty} (L^+\mu)(x_n) \lim_{n\to\infty} \int_\Omega u_y(x_n) d\mu(y) = \int_\Omega u_y(x) d\mu(y) = (L^+\mu)(x).$$

The boundedness of the map now follows from the inequality

$$\left| \int_\Omega u_y(x) d\mu \right| \leq \int_\Omega |u_y(x)| \, d\,|\mu| \leq \sup_{y\in\Omega} \|u_y\| \, \|\mu\|_{TV} \,, \; x \in \Omega.$$

Now we show that $P(D)L^+\mu = \mu$. For this, let $\phi \in \mathcal{C}_c^\infty(\Omega)$ be arbitrary. We then have

$$\int_\Omega (L^+\mu)(y) P(D_y)^* \phi(y) dy = \int_\Omega \int_\Omega u_x(y) P(D_y)^* \phi(y) d\mu(x) dy,$$

where $P(D)^*$ denotes the adjoint to $P(D)$. The function $(x, y) \mapsto u_x(y) P(D_y)^* \phi(y)$ is continuous and supported on a set of the form $\Omega \times C$, where C is compact. As such, it is integrable with respect to the measure $\mu \otimes dy$, and we may apply Fubini's theorem. Subsequently shifting $P(D_y)$ onto u_x and utilizing $P(D_y)u_x = \delta_x$, we obtain that the above is equal to

$$\int_\Omega \phi(x) d\mu(x).$$

This exactly means that $P(D)L^+\mu = \mu$. □

Now we may prove Lemma 1 about the properties of $\mathcal{B}_P$ as a normed space.

Proof of Lemma 1. The only non-trivial step in proving that $\|u\|_{\mathcal{B}_P}$ is a norm is to prove that $\|u\|_P = 0 \Rightarrow u = 0$. This follows from the assumption on the set K: If $\|u\|_{B_P} = 0$, then in particular $P(D)u = 0$ and $u = 0$ in K. Since $u = 0$ is a function obeying $P(D)u = 0$ and $u = 0$ in K, the uniqueness of the continuation implies that u must vanish everywhere in Ω.

To prove that $\mathcal{B}_P$ is a Banach space, notice that we can interpret $\mathcal{B}_P$ as a subspace of the Banach space $\mathcal{M}(\Omega) \times \mathcal{M}(K)$. This space is furthermore closed: If $(P(D)u_n, u_n) \to (\mu, \overline{u})$ in $\mathcal{M}(\Omega) \times \mathcal{M}(K)$, there must be $P(D)\overline{u} = \mu$ on K.

To see this, let $\phi \in \mathcal{C}_0(K)^\infty$ be arbitrary. We then have $P(D)^*\phi \in \mathcal{C}_0(K)$, and consequently

$$\int_K \overline{u} P(D)^* \phi dx = \lim_{n\to\infty} \int_K u_n P(D)^* \phi dx = \lim_{n\to\infty} \int_K \phi d(P(D)u_n) = \int_K \phi d\mu,$$

where we used the fact $P(D)u_n \to \mu$ in the last step. Since $P(D)L^+\mu = \mu$, we conclude that $P(D)(\overline{u} - L^+\mu) = 0$ in K. The continuation property implies that there exists a $\widehat{u}$ with $P(D)\widehat{u} = 0$ in Ω and $\widehat{u} = \overline{u} - L^+\mu$ in K. We then have $\mu = P(D)(\widehat{u} + L^+\mu)$ and $(\widehat{u} + L^+\mu)|_K = \overline{u}$, so that $(\mu, \overline{u}) \in \mathcal{B}_P$. □

Now let us prove Proposition 3

Proof of Proposition 3. In Lemma 9, we showed that $P(D)L^+ = \mathrm{Id}_{\mathcal{M}}$. This already proves that $\operatorname{ran} P(D) = \mathcal{M}(\Omega)$. Also, it shows that L^+ is a continuous operator from $\mathcal{M}(\Omega)$ to $\mathcal{B}_P(\Omega)$: $\mathcal{C}(\Omega) \hookrightarrow \mathcal{M}(K)$ due to the boundedness of K, and if $\mu_n \to \mu$ in $\mathcal{B}_P$, then $P(D)L^+\mu_n = \mu_n \to \mu = P(D)L^+\mu$.

It follows that $L^+P(D) = \mathrm{Id}_{\operatorname{ran} L^+}$. If we can prove that $\operatorname{ran} L^+$ is closed, we have shown that $\ker P(D)$ has the closed complementary subspace $\operatorname{ran} L^+$.

To show the latter, let $u_n = L^+\mu_n$ in $\operatorname{ran} L^+$ converge to an element $\overline{u} \in \mathcal{B}_P$. Then, by definition of $\mathcal{B}_P$, $P(D)L^+\mu_n = \mu_n \to P(D)\overline{u}$. Consequently, by the continuity of L^+,

$$u_n = L^+\mu_n = L^+P(D)L^+\mu_n \to L^+P(D)\overline{u},$$

so that $\overline{u} = L^+P(D)\overline{u} \in \operatorname{ran} L^+$.

It remains to calculate the operator $(L^+)^*$. For $a \in L^1(\Omega)$ and $\mu \in \mathcal{M}(\Omega)$, we have

$$\left\langle (L^+)^* a, \mu \right\rangle = \left\langle a, L^+\mu \right\rangle = \int_\Omega \int_\Omega a(y) u_x(y) d\mu(x) dy$$

$(x, y) \to u_x(y)a(y)$ is in $L^1(\mu \otimes dy)$, so that we may apply Fubini and obtain

$$\left\langle (L^+)^* a, \mu \right\rangle \int_\Omega \left(\int_\Omega a(y) u_x(y) dy \right) d\mu(x).$$

The last assertion about $(L^+)^* a \in \mathcal{C}_0(\Omega)$ is argued as follows. Let $\epsilon > 0$. First, since $a \in L^1(\Omega)$, there exists a compact set C such that $\left\| a_{\Omega\setminus C} \right\|_1 \leq \epsilon$. Further, since the map $x \mapsto u_x$ is vanishing at infinity as a map from Ω to $\mathcal{C}(\Omega \cap B_R(0))$, there exists a compact set $\widetilde{C}$ such that if $x \notin \widetilde{C}$, $\|u_x\|_{\mathcal{C}(C)} \leq \epsilon$. This implies for such x

$$\begin{aligned} \left| (L^+)^* a(x) \right| &\leq \|u_x\|_\infty \int_{\Omega\setminus C} |a(y)|\, dy + \|u_x\|_{\mathcal{C}(C)} \int_C |a(y)|\, dy \\ &\leq \|u_x\|_\infty \epsilon + \epsilon \|a\|_1 , \end{aligned}$$

so that the theorem is proved. □

Now let us finally argue that the differential operators of the form (9) can be included in our framework.

Proof of Proposition 4. Consider the space $\mathcal{H}_0^k(\Omega)$, defined as the closure of $\mathcal{C}_0^\infty(\Omega)$ in the Sobolev norm $\|\cdot\|_{H^k(\Omega)}$. We can formulate the problem $P(D)u = f$ as an operator equation on $\mathcal{H}_0^k(\Omega)$ as follows:

$$\langle P(D)u, v\rangle = \int_\Omega \sum_{|\alpha|=k} \sum_{|\beta|=k} p_{\alpha,\beta}(x) D^\alpha u(x) D^\beta(x) dx = \langle f, v\rangle, v \in \mathcal{H}_0^k(\Omega).$$

By the Lax-Milgram lemma together with the ellipticity condition, this problem has a unique solution as soon as $f \in \mathcal{H}_0^k(\Omega)$. Now, since $k > d/2$, we have the continuous Sobolev embedding $\mathcal{H}_0^k(\Omega) \hookrightarrow \mathcal{C}_0(\Omega)$. (For $\Omega = \mathbb{R}^d$, this can be proven with Fourier methods, for a bounded domain, this is a Sobolev embedding theorem.) This both proves that $\delta_x \in \mathcal{H}_0^k(\Omega)^*$ and that the solution $u_x \in \mathcal{C}_0(\Omega)$.

To show that the map $x \to u_x$ is vanishing at infinity on compact sets, let us first assume that Ω is bounded. When x escapes to infinity, $\delta_x \overset{*}{\rightharpoonup} 0$ in $\mathcal{M}(\Omega)$, and therefore also in $\mathcal{H}_0^k(\Omega)$. The "continuous dependence on the data"-part of Lax-Milgram theorem therefore implies that $u_x \rightharpoonup 0$ in $\mathcal{H}_0^k(\Omega)$. Since the embedding $\mathcal{H}_0^k(\Omega) \hookrightarrow \mathcal{C}_0(\Omega)$ in this case even is compact (see e.g. [1, Theorem 6.2]), this implies that $u_x \to 0$ in $\mathcal{C}_0(\Omega)$, which was to be proven.

Now let $\Omega = \mathbb{R}^d$ and $R > 0$ be arbitrary. The result [20, Theorem 10.2.1] states that the solution u_x is equal to $\Phi(\cdot - x)$ for a Φ obeying

$$\sup_{\xi \in \mathbb{R}^d} \left| \widetilde{P}(\xi) \widehat{\Phi}(\xi) \right| < \infty,$$

where $\widetilde{P}$ is defined as

$$\widetilde{P} = \left(\sum_{\alpha \geq 0} |D^\alpha P(\xi)|^2 \right)^{1/2}.$$

By using the ellipticity assumption, one sees that this implies that $(1+|\xi|^{2k})|\hat{\Psi}| \leq C$, which ensures that $\left|\hat{\Psi}\right|$ is integrable ($k > d/2$). By the Riemann-Lebesgue theorem, $\Psi \in \mathcal{C}_0(\mathbb{R}^d)$. This already implies that $u_x = \Psi(\cdot - x)$ vanishes to infinity on compact sets.

To prove the final claim, let u obey $P(D)u = 0$ on K. Then in particular $u \in \mathcal{H}^k(K)$. This implies that for every set $K \subseteq \widetilde{\Omega} \subseteq \Omega$ with $\operatorname{dist}(K, \partial\Omega) > 0$, there exist a function $\tilde{u} \in \mathcal{H}^k(\mathbb{R}^d)$ with compact support in $\widetilde{\Omega}$ (see [1, Theorem 2.8]) such that $u = \widetilde{u}$ on K. Now consider the following problem:

$$\begin{cases} P(D)u = -P(D)\widetilde{u}, \ x \in \Omega \backslash K \\ D^\alpha u|_{\partial K \cup \partial\Omega} = 0 \end{cases}.$$

This problem can be shown to have a solution $\widehat{u}$. Now consider the function

$$\overline{u}(x) = \begin{cases} u(x), x \in K \\ -\widehat{u}(x) + \widetilde{u}(x), x \in \Omega \notin K \end{cases}$$

Due to boundary term cancellation, together with the fact that the $\overline{u}_K$ and $\overline{u}_{\Omega \notin K}$ solves the problem $P(D)u$ in their respective domains, this function obeys $P(D)\overline{u} = 0$ in Ω, and of course $\overline{u}|_K = u$.

As for the uniqueness of the extension, we note that if $P(D)u = 0$ on Ω, the ellipticity assumption implies that $\sum_{\alpha=k} |D^\alpha u|_2^2 = 0$. This in particular implies that $\Delta^k u = 0$, i.e., $\Delta^{k-1}u$ is harmonic. Since $\Delta^{k-1}u$ vanishes on K, and K has non-empty interior, it must vanish everywhere (this is the identity theorem of harmonic functions). By repeating this argument k times, we finally obtain that u vanishes on Ω. □

4.4 Miscellaneous

Here, the rest of the left out proofs are given. We start with the simple proposition about existence of solutions to the problem (1).

Proof of Proposition 1. Let $(u_n)_{n\in\mathbb{N}}$ be a minimizing sequence for (1). Let us write $u_n = L^+\mu_n + u_{n,K}$ with $\mu_n \in \mathcal{M}(\Omega)$ and $u_{n,K} \in \ker L$ for each $n \in \mathbb{N}$. We may thereby without loss of generality assume that $u_{n,K} \in \operatorname{span}(w_\ell)_{\ell=1}^{\widehat{m}}$, where w_ℓ are vectors such that $(Aw_\ell)_{\ell=1}^{\widehat{m}}$ spans $A(\ker L)$ (any alteration of $u_{K,n}$ not parallel to this space will neither change the value of $\|Lu\|$ or the value of $f_b(Au)$.

Now, due to the minimization property of the sequence,

$$(\mu_n)_{n\in\mathbb{N}} \text{ and } (f_b(A(L^+\mu_* + u_{n,K}))_{n\in\mathbb{N}}$$

are both bounded. Due to the coercivity of f_b together with the fact that A restricted to the space $(w_\ell)_{\ell=1}^{\widehat{m}}$ is injective, the sequence $(u_{n,K})_{n\in\mathbb{N}}$ will be bounded in $A(\ker L)$. Due to the Banach-Alaoglu theorem and the separability of $\mathcal{C}_0(\Omega)$ (i.e. the pre-dual of $\mathcal{M}(\Omega)$), $(\mu_n)_{TV}$ will contain a subsequence which converges to, say, μ^*. Similarly, since $(u_{n,K})$ lives in the finite-dimensional space $\operatorname{span}(w_\ell)_{\ell=1}^{\widehat{m}}$, it will also contain a subsequence convergent to, say u^*. Now, using the same notation for the convergent subsequences as for the sequences themselves, we have

$$\begin{aligned}\|\mu^*\|_{TV} + f_b(A(L^+\mu^* + u_K^*)) &\leq \liminf \|\mu_n\|_{TV} + f_b(A(L^+\mu_n + u_{n,K})) \\ &= \liminf \|Lu_n\|_{TV} + f_b(Au_n) = \min_{u\in\mathcal{B}} \|Lu\|_{TV} + f_b(Au).\end{aligned}$$

We used Lemma 7 and the lower semicontinuity of f_b and of the TV-norm. Hence, $L^+\mu^* + u_K^*$ is the solution whose existence we had to prove. □

Now let us include spline-admissible operators in our framework.

Proof of Lemma 2. 1. The finite-dimensionality of $\ker L$ is simply assumption 3 of Theorem 1 of [30]. Theorem 4 and 5 of [30] proves that L has a right inverse L_Φ^{-1}. This implies that

$$\operatorname{ran} L \subseteq \operatorname{ran} LL_\Phi^{-1} = \operatorname{ran} \mathrm{Id} = \mathcal{M}.$$

2. The space $\mathcal{C}_L$ as defined in Theorem 6 of [30] is defined as

$$\mathcal{C}_L = L^*(\mathcal{C}_0(\mathbb{R}^d)) + \operatorname{span}(\phi_i)_{i=1}^r,$$

where ϕ_i is a system of functionals which restricted to $\ker L$ becomes a of the dual of $\ker L$. Without loss of generality, we can assume that $\phi_i|_V = 0$ for each i (if not, we could instead consider the operators $\widetilde{\phi}_i = \phi_i \Pi_{\ker L}$).

Then if $a \in \mathcal{C}_L$, we have

$$(L^+)^* a = (L^+)^* L^* \rho + \sum_{i=1}^{r} \gamma_i (L^+)^* \phi_i$$

for some $\rho \in \mathcal{C}_0(\mathbb{R}^d)$ and γ_i. Now $(L^+)^* L^*) = (LL^+)^* = \Pi^*_{\operatorname{ran} L} = \mathrm{Id}$ and $(L^+)^* \phi_i = 0$, so that $(L^+)^* a = \rho \in C_0(\mathbb{R}^d)$.

If on the other $(L^+)^* a \in \mathcal{C}_0(\mathbb{R}^d)$, we have

$$L^* \mathcal{C}_0(\mathbb{R}^d) \ni L^* (L^+)^* a = (L^+ L)^* a = \Pi^*_V a.$$

Since each functional $a \in \mathcal{M}_L^*$ can be written as $\Pi^*_V a + \Pi^*_{\ker L} a$, and $\Pi^*_{\ker L} a \in \operatorname{span}(\phi_i)_{i=1}^r$, $a \in \mathbb{C}_L$. □

Next, we discuss the case of L being the differential operator on $BV((0,1)$.

Proof of Lemma 2. Note that we have $\ker L = \operatorname{span}(1)$, the vector space of constant functions on Ω, hence the space V can be identified with the space of functions with zero mean:

$$V = \left\{ u \in BV(\Omega), \int_\Omega u(t)\, dt = 0 \right\}.$$

For $\mu \in \mathcal{M}$, consider the mapping $I : \mu \mapsto u$ defined for $s \in [0,1]$ by $u(s) = \mu([0,s])$. We only need to prove that $DI(\mu) = \mu$ in the distributional sense. Let $\phi \in C_c^\infty(\Omega)$:

$$\begin{aligned}
\langle I(\mu), \phi' \rangle &= \int_0^1 \mu([0,t]) \phi'(t)\, dt \\
&= \int_0^1 \int_0^1 1_{[0,t]}(s) d\mu(s) \phi'(t)\, dt \\
&= \int_0^1 \int_0^1 1_{[s,1]}(t) \phi'(t)\, dt d\mu(s) \\
&= \int_0^1 -\phi(s) d\mu(s) = -\langle \mu, \phi \rangle.
\end{aligned}$$

This proves the surjectivity of L. We see that the proposed form of L^+ is the right one, since $s \mapsto \mu([0,s]) - \int_0^1 \mu([0,s]) ds$ is a function of zero mean.

We now calculate

$$\begin{aligned}
\left\langle (L^+)^* \xi, \mu \right\rangle &= \left\langle \xi, L^+ \mu \right\rangle \\
&= \int_0^1 \xi(t) \left(\int_0^1 1_{[0,t]}(s) d\mu(s) - \int_0^1 \mu([0,r]) dr \right) dt \\
&= \int_0^1 \left(\int_0^1 1_{[s,1]}(t) \xi(t) dt \right) d\mu(s) - \int_0^1 \xi(t) dt \cdot \int_0^1 1_{[0,r]}(s) d\mu(s) dr \\
&= \int_0^1 \left(\int_s^1 \xi(t) dt \right) d\mu(s) - \int_0^1 \xi(t) dt \cdot \int_0^1 (1-s) d\mu(s)
\end{aligned}$$

In particular, the action of $(L^+)^* \xi$ is given by a continuous function, which is vanishing on the boundary of $(0,1)$

□

5 Conclusion & Outlook

In this paper we have studied the properties of total variation regularized problems, where total-variation should be understood as a term of form $\|Lu\|_{TV}$, with L a linear operator. We have shown that under a convexity assumption on the data-fit term, some of the solutions $\hat{u}$ of total-variation regularized inverse problems are m-sparse, where m denotes the number of measurements. This precisely means that $L\hat{u}$ is an atomic measure supported on at most m points. This result extends recent advances [30], by relaxing some hypotheses on the linear operator L and on the domain of the functions.

The second contribution of this paper is to show that solutions of this infinite dimensional problem can be obtained by solving one or two consecutive finite dimensional problems, given that the measurements belong to some function spaces such as the trigonometric polynomials or the set of piecewise linear functions on polyhedral domains. Once again, this result extends significantly recent results on super-resolution [8, 26]. The analysis provided for piecewise linear functions is novel and we believe that it might have important consequences in the numerical analysis of infinite dimensional inverse problems: the scaling with respect to the number of grid points is just linear, contrarily to approaches based on semi-definite relaxations or Lasserre hierarchies.

As an outlook, we want to stress out that the hypotheses formulated on the linear operator L rule out a number of interesting applications, such as total variation regularization in image processing. We plan to study how the results and the proof techniques in this paper could apply to more general cases.

Acknowledgement

A. Flinth acknowledges support from the Deutsche Forschungsgemeinschaft (DFG) Grant KU 1446/18-1, and from the Berlin Mathematical School (BMS). He also wishes to thank Yann Traonmillin, Felix Voigtländer and Philipp Petersen for interesting discussions on this subject. This work was partially funded by ANR JCJC OMS. P. Weiss wishes to thank Michael Unser warmly for motivating him to work on the subject at the second OSA "Mathematics in Imaging" conference in San Francisco and for providing some insights on his recent paper [30]. In addition, he thanks Didier Henrion particularly and Alban Gossard, Frédéric de Gournay, Jonas Kahn, Etienne de Klerk, Jean-Bernard Lasserre, Michael Overton and Lieven Vandenberghe for interesting feedbacks and insights on a preliminary version of this work. The two authors wish to thank Gitta Kutyniok from TU Berlin for supporting this research.

References

[1] R. Adams. *Sobolev Spaces*. Academic Press, 1975.

[2] B. Adcock and A. C. Hansen. Generalized sampling and infinite-dimensional compressed sensing. *Found. of Comp. Math.*, 16(5):1263–1323, 2016.

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

[4] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*, volume 254. Clarendon Press Oxford, 2000.

[5] J. Borwein and A. Lewis. Partially finite convex programming, part I: Quasi relative interiors and duality theory. *Math. Prog.*, 57(15):15–48, 1992. doi:10.1007/BF01581072.

[6] A. Braides. *Gamma-convergence for Beginners*, volume 22. Clarendon Press, 2002.

[7] K. Bredies and H. K. Pikkarainen. Inverse problems in spaces of measures. *ESAIM: Contr Optim Ca.*, 19(1):190–218, 2013.

[8] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Commun. Pur. Appl. Math.*, 67(6):906–956, 2014.

[9] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE T. Inform. Theory*, 52(2):489–509, 2006.

[10] A. Chambolle, V. Duval, G. Peyré, and C. Poon. Geometric properties of solutions to the total variation denoising problem. *Inv. Probl.*, 33(1):015002, 2016.

[11] S. Chandrasekaran, M. Gu, J. Xia, and J. Zhu. A fast QR algorithm for companion matrices. *Oper. Th. Adv. A*, 179:111–143, 2007.

[12] Y. De Castro, F. Gamboa, D. Henrion, and J.-B. Lasserre. Exact solutions to super resolution on semi-algebraic domains in higher dimensions. *IEEE T. Inform. Theory*, 63(1):621–630, 2017.

[13] D. L. Donoho. Compressed sensing. *IEEE T. Inf. Theory*, 52(4):1289–1306, 2006.

[14] B. Dumitrescu. *Positive trigonometric polynomials and signal processing applications*, volume 103. Springer, 2007.

[15] V. Duval and G. Peyré. Exact support recovery for sparse spikes deconvolution. *Found. Comp. Math.*, 15(5):1315–1355, 2015.

[16] I. Ekeland and R. Temam. *Convex analysis and variational problems*. SIAM, 1999.

[17] M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inv. Probl.*, 23(3):947, 2007.

[18] S. Fisher and J. Jerome. Spline solutions to l1 extremal problems in one and several variables. *J. Approx. Theory.*, 13(1):73–83, 1975.

[19] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*, volume 1. Birkhäuser Basel, 2013.

[20] L. Hörmander. *The analysis of Linear Partial Differential operators II.* Springer, 1963.

[21] B. Kaltenbacher, A. Neubauer, and O. Scherzer. *Iterative regularization methods for nonlinear ill-posed problems*, volume 6. Walter de Gruyter, 2008.

[22] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. on Optimiz.*, 11(3):796–817, 2001.

[23] M. MGrant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, mar 2014.

[24] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 1992.

[25] W. Rudin. *Functional analysis.* International series in pure and applied mathematics. McGraw-Hill, Inc., New York, 1991.

[26] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. *IEEE T. Inform. Theory*, 59(11):7465–7490, 2013.

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

[28] Y. Traonmilin, G. Puy, R. Gribonval, and M. Davies. Compressed sensing in Hilbert spaces. *arXiv preprint arXiv:1702.04917*, 2017.

[29] M. Unser, J. Fageot, and H. Gupta. Representer Theorems for Sparsity-Promoting ℓ_1 Regularization. *IEEE T. Inform. Theory*, 62(9):5167–5180, 2016.

[30] M. Unser, J. Fageot, and J. P. Ward. Splines are universal solutions of linear inverse problems with generalized-TV regularization. *arXiv preprint arXiv:1603.01427*, 2016.

[31] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM review*, 38(1):49–95, 1996.