



Citation for published version:

Bredies, K, Carioni, M, Holler, M, Korolev, Y & Schönlieb, C-B 2023 'A sparse optimization approach to infinite infimal convolution regularization' arXiv.

Publication date:
2023

[Link to publication](#)

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

A sparse optimization approach to infinite infimal convolution regularization

Kristian Bredies* Marcello Carioni[†] Martin Holler[‡] Yury Korolev[§]
 Carola-Bibiane Schönlieb[¶]

Abstract

In this paper we introduce the class of *infinite infimal convolution functionals* and apply these functionals to the regularization of ill-posed inverse problems. The proposed regularization involves an infimal convolution of a continuously parametrized family of convex, positively one-homogeneous functionals defined on a common Banach space X . We show that, under mild assumptions, this functional admits an equivalent convex lifting in the space of measures with values in X . This reformulation allows us to prove well-posedness of a Tikhonov regularized inverse problem and opens the door to a sparse analysis of the solutions. In the case of finite-dimensional measurements we prove a *representer* theorem, showing that there exists a solution of the inverse problem that is sparse, in the sense that it can be represented as a linear combination of the extremal points of the ball of the lifted infinite infimal convolution functional. Then, we design a generalized conditional gradient method for computing solutions of the inverse problem without relying on an a priori discretization of the parameter space and of the Banach space X . The iterates are constructed as linear combinations of the extremal points of the lifted infinite infimal convolution functional. We prove a sublinear rate of convergence for our algorithm and apply it to denoising of signals and images using, as regularizer, infinite infimal convolutions of fractional-Laplacian-type operators with adaptive orders of smoothness and anisotropies.

2020 Mathematics Subject Classification: 65J20, 65K10, 49J45, 35R11.

Keywords: infimal convolution, sparse optimization, fractional Laplacian regularization, learning regularizers, conditional gradient methods

Contents

| | | |
|----------|---|----------|
| 1 | Introduction | 2 |
| 2 | The infinite infimal convolution functional | 5 |
| 2.1 | Infinite infimal convolution | 6 |
| 2.2 | Tikhonov regularized inverse problems | 7 |
| 3 | Convex lifting | 7 |
| 3.1 | Convex lifting of the regularizer | 7 |
| 3.2 | Convex lifting of the variational problem | 11 |
| 3.3 | Existence of minimizers | 12 |
| 3.4 | Extremal points of the lifted regularizer | 13 |

*Kristian Bredies, Institute of Mathematics and Scientific Computing, University of Graz, Heinrichstraße 36, A-8010 Graz, Austria. Email: kristian.bredies@uni-graz.at

[†]Marcello Carioni, Department of Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands. Email: m.c.carioni@utwente.nl

[‡]Martin Holler, Institute of Mathematics and Scientific Computing, University of Graz, Heinrichstraße 36, A-8010 Graz, Austria. Email: martin.holler@uni-graz.at

[§]Yury Korolev, Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK. Email: y mk30@bath.ac.uk

[¶]Carola-Bibiane Schönlieb. Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, CB3 0WA, Cambridge, UK. Email: cbs31@cam.ac.uk

| | | |
|----------|---|-----------|
| 4 | A representer theorem for the infinite infimal convolution functional | 15 |
| 5 | A generalized conditional gradient method | 16 |
| 5.1 | Description of the algorithm | 17 |
| 5.1.1 | Insertion step | 18 |
| 5.1.2 | Weights optimization step | 18 |
| 5.1.3 | Generalized conditional gradient method for the lifted problem | 19 |
| 5.1.4 | Generalized conditional gradient method for the infinite infimal convolution functional | 20 |
| 5.2 | Sublinear convergence | 21 |
| 6 | Illustrative applications and numerics | 23 |
| 6.1 | Learning the orders of 1-dimensional fractional Laplacians in an infinite infimal convolution | 23 |
| 6.1.1 | Setting | 24 |
| 6.1.2 | The generalized conditional gradient method | 25 |
| 6.1.3 | Numerical experiments | 26 |
| 6.2 | Learning the anisotropies of fractional Laplacians in an infinite infimal convolution | 28 |
| 6.2.1 | Setting | 28 |
| 6.2.2 | The generalized conditional gradient method | 29 |
| 6.2.3 | Numerical experiments | 29 |
| 7 | Conclusions | 33 |
| 8 | Acknowledgments | 33 |
| A | Measures with values in Banach spaces | 36 |
| A.1 | L^p -functions with values in Banach spaces | 36 |
| A.2 | Definition and Radon–Nikodým property | 36 |
| A.3 | Duality theory | 38 |
| B | Fractional-Laplacian-type operators | 39 |
| B.1 | Periodic L^2 -functions | 39 |
| B.2 | Fourier series of periodic L^2 -functions | 39 |
| B.3 | Periodic fractional Sobolev functions | 39 |
| B.4 | The fractional Laplacian | 40 |
| B.5 | Extremal points of L^2 -balls of fractional-Laplacian-type operators | 40 |

1 Introduction

The infimal convolution of two convex functions $f_1, f_2 : X \rightarrow (-\infty, +\infty]$ is a classical convexity-preserving operation defined as follows

$$(f_1 \square f_2)(x) = \inf_{x_1+x_2=x} f_1(x_1) + f_2(x_2). \quad (1.1)$$

Many optimization problems can be formulated using this notion: not only in economics, where infimal convolutions are related to the so-called Pareto optimality [31, 47, 50], but also in other fields such as engineering, physics, social sciences, and mechanics [4]. More recently, infimal convolutions of convex functionals have been used to regularize inverse problems with applications to image and signal processing. One of the first successful approaches in this direction has been proposed in [22] in the context of image denoising. It is well-known that reconstructions obtained using variational models that penalize the L^1 -norm of the gradient of the image, such as the ROF model [42], achieve good performances to reconstruct sharp edges, but exhibit artifacts in the smooth regions of the

image (the so-called staircasing effect). On the other hand, higher-order regularizers tend to smooth out the edges but perform better in the regions where a small gradient is expected. Following these considerations, the authors in [22] introduced a regularizer defined as the infimal convolution of a first and a second order functional as follows

$$R(v) = \inf_{v_1+v_2=v} \alpha \int_{\Omega} |\nabla v_1| dx + \beta \int_{\Omega} |\nabla(\nabla v_2)| dx, \quad (1.2)$$

where $\alpha, \beta > 0$ are fixed parameters, with the goal of automatically selecting the regions of the image where it is convenient to apply a first order regularizer and the regions where the second order regularizer is more suitable. In the following years, further regularizers defined as the infimal convolution of functionals were introduced, achieving striking results in solving inverse problems. Combinations of higher-order functionals were explored improving on the work of [22], see [8, 23, 45], and spatio-temporal infimal convolutions were used for solving dynamic inverse problems [32, 44]. Other more recent examples can be found in [14, 17–20, 29, 30, 35].

In this work we focus on a generalization of the infimal convolution functional that is performed on an infinite (and even uncountable) collection of convex, non-negative, positively one-homogeneous, and coercive functionals $J(\cdot, s)$ defined on a common dual Banach space X and parametrized by an index $s \in S$, where S is a compact set in the Euclidean space. We additionally suppose that J is lower semicontinuous in $X \times S$, where X is endowed with the weak* topology. Given a positive measure $\sigma \in \mathcal{M}_+(S)$ that models which convex functional is “active” in the infimal convolution, we define the operator $R : X \times \mathcal{M}_+(S) \rightarrow [0, \infty]$ as follows

$$R(v, \sigma) := \inf \left\{ \int_S J(u(s), s) d\sigma(s) : u \in L^1((S, \sigma); X) \text{ with } \int_S u(s) d\sigma(s) = v \right\},$$

where X is a given Banach space and $L^1((S, \sigma); X)$ is the Bochner space of L^1 functions defined on (S, σ) with values in X . Following [41], we call R the *infinite infimal convolution functional* to stress the fact that we are allowing for an infinite family of convex functionals continuously parametrized by $s \in S$ and we search for the optimal decomposition of v with respect to σ , i.e., $\int_S u(s) d\sigma(s) = v$, among all $u \in L^1((S, \sigma); X)$. We remark that, since each $J(\cdot, s)$ is positively one-homogeneous, if σ is chosen as a finite sum of Dirac deltas concentrated in the points $s_1, \dots, s_N \in S$, then $R(v, \sigma)$ is the standard infimal convolution of the functionals $J(\cdot, s_1), \dots, J(\cdot, s_N)$. Our aim is to use $R(v, \sigma)$ as a regularizer for inverse problems $Av = f$, where $A : X \rightarrow Y$ is a linear measurement operator and $f \in Y$ is the observation. In particular, we solve the Tikhonov regularized problem

$$\inf_{\substack{v \in X, \\ \sigma \in \mathcal{M}_+(S)}} \frac{1}{2} \|Av - f\|_Y^2 + \alpha R(v, \sigma), \quad (1.3)$$

where $\alpha > 0$ is a positive parameter. Note that in (1.3) we are simultaneously optimizing over $v \in X$ and $\sigma \in \mathcal{M}_+(S)$. In this way, (1.3) is automatically selecting the best combination of functionals $J(\cdot, s)$ that optimally regularizes the reconstruction of a given measurement f . Our setting is presented in more detail in Section 2.

We start our analysis by proving well-posedness of (1.3), which is done in Section 3. Compared to the classical infimal convolution functional, this is not straightforward. Indeed, since the Bochner space $L^1((S, \sigma); X)$ does not admit a predual and the functionals $J(\cdot, s)$ are only assumed to be coercive, standard methods based on the Banach–Alaoglu theorem cannot be applied. To circumvent this obstacle we leverage the positive one-homogeneity of $J(\cdot, s)$ to construct a convex lifting of the infinite infimal convolution functional to the space of Radon measures with values in X , denoted by $\mathcal{M}(S; X)$. Formally, such convex lifting is obtained by setting $\mu = u\sigma \in \mathcal{M}(S; X)$ defined as the Bochner integration of $u \in L^1((S, \sigma); X)$ against the measure $\sigma \in \mathcal{M}_+(S)$. We define the convex lifting of the functional $F_{\sigma}(u) := \int_S J(u(s), s) d\sigma(s)$ as $\hat{F} : \mathcal{M}(S; X) \rightarrow [0, +\infty]$ defined by

$$\hat{F}(\mu) = \int_S J\left(\frac{\mu}{|\mu|}(s), s\right) d|\mu|(s), \quad (1.4)$$

where $|\mu| \in \mathcal{M}_+(S)$ is the variation (measure) of $\mu \in \mathcal{M}(S; X)$ defined for every Borel set $E \subset S$ as

$$|\mu|(E) = \sup \left\{ \sum_{i=1}^n \|\mu(A_i)\|_X : n \in \mathbb{N}, (A_i)_{i=1}^n \text{ partition of } E \right\}$$

and $\frac{\mu}{|\mu|} \in L^1((S, \sigma); X)$ is the density of μ with respect to $|\mu|$ defined by the relation $\mu(E) = \int_E \frac{\mu}{|\mu|}(s) d|\mu|(s)$ for every Borel set $E \subset S$. We then rewrite the lifted inverse problem as

$$\inf_{\mu \in \mathcal{M}(S; X)} \frac{1}{2} \|A\mu(S) - f\|_Y^2 + \alpha \hat{F}(\mu). \quad (1.5)$$

Relying on the positive one-homogeneity of $J(\cdot, s)$ we show that (1.5) and (1.3) are equivalent. Moreover, after proving that \hat{F} is coercive and weakly-* lower semicontinuous, we obtain the well-posedness of (1.5) and (1.3) using standard arguments of calculus of variations.

The convex lifting of the infinite infimal convolution functional is not only useful to show well-posedness of the original problem, but it also sheds light on the sparsity properties of the regularized inverse problem. The key observation is that the extremal points of the unit ball of \hat{F} , denoted by B , can be characterized as follows:

- If $0 \in \text{Ext}(B_s)$ for each $s \in S$, then $\text{Ext}(B) = \{v\delta_s : s \in S, v \in \text{Ext}(B_s)\}$.
- If there exists $s \in S$ such that $0 \notin \text{Ext}(B_s)$, then $\text{Ext}(B) = \{v\delta_s : s \in S, v \in \text{Ext}(B_s) \setminus \{0\}\}$,

where B_s is the unit ball of $J(\cdot, s)$ and $\text{Ext}(B_s)$ is the set of its extremal points. As shown in [10] and [7], the extremal points of the unit ball of the regularizer provide information on the sparse structure of the minimizers. In particular, such extremal points are precisely the atoms used to represent sparse solutions in case of finite-dimensional data. Such results are called representer theorems and have recently become relevant in the context of inverse problems [7, 10], optimization [49], and machine learning [6, 48]. Relying on the result of [10] and the characterization of the extremal points of the lifted regularizer, we prove a representer theorem for inverse problems regularized with the infimal convolution functional in Section 4. In particular, we show that if the data is Q -dimensional for $Q \in \mathbb{N}$, then a solution (v^*, σ^*) of (1.3) can be written as follows

$$v^* = \sum_{i=1}^M \lambda_i v_i, \quad \sigma^* = \sum_{s \in \{s_i: i=1, \dots, M\}} \left\| \sum_{j \in \{1, \dots, M: s_j=s\}} \lambda_j v_j \right\|_X \delta_s, \quad (1.6)$$

where $\lambda_i > 0$, $s_i \in S$, $v_i \in \text{Ext}(B_{s_i})$, and $M \leq Q$.

In Section 5 we leverage the knowledge of the atoms of the lifted regularization functional to design a generalized conditional gradient method for solving (1.3). Generalized conditional gradient methods (GCG) are infinite-dimensional generalizations of the classical Frank–Wolfe algorithms [28]. In the context of inverse problems they have been used to design algorithms that do not require an a priori discretization of the domain X (also called off-the-grid algorithms), since optimization steps are performed by inserting iteratively suitably chosen elements of X . Some applications of generalized conditional gradient methods to inverse problems can be found, for example, in [15, 16]. More recently, GCG methods have been studied as sparse optimization algorithms in infinite-dimensional spaces. Indeed, it has been shown that it is possible to construct the k th iterate of a GCG method as the following sparse element

$$u^k = \sum_{i=1}^{N_k} \lambda_i^k u_i^k,$$

where $N_k \in \mathbb{N}$, $\lambda_i^k > 0$ and u_i^k are extremal points of the unit ball of the convex regularizer. Such observation has proved useful, for example, in solving super-resolution problems in the space of measures [9, 27, 39] and in dynamic inverse problems with regularized with optimal transport

energies [12, 26]. In this paper we use the characterization of the extremal points of the unit ball of \hat{F} given in Section 3 to design a GCG for the lifted problem (1.5). We follow the general procedure described in [13] and adapt the results of [13] to prove sublinear convergence of the GCG algorithm. Moreover, we use the equivalence of (1.3) and (1.5) to rewrite the GCG algorithm for (1.5) as a converging algorithm for the inverse problem regularized with the infinite infimal convolution functional (1.3) with the same rate of convergence.

Finally, in Section 6 we provide examples of signal and image denoising problems regularized with an infinite infimal convolution functional, where we apply our generalized conditional gradient method. First we consider the task of denoising a signal using infimal convolutions of L^2 -norms of fractional-Laplacian-type operators with adaptive order $s \in [0, 1]$. We refer to [3, 5] for examples of applications of fractional-Laplacian regularizers to image denoising. We construct the regularizer in such a way that the infimal convolution is learning the predominant frequencies of the data and is assigning a lower-order regularization to them. As a second application we consider denoising images using infinite infimal convolutions of L^2 -norms of anisotropic fractional-Laplacian-type operators. Here the anisotropy direction is regulated by a parameter $s \in [0, \pi]$ providing the direction $(\cos(s), \sin(s)) \in \mathbb{S}^1$. The infinite infimal convolution regularizer learns the directions of higher oscillation of the noisy image f and applies a lower regularization in such directions. In this way the natural anisotropy of the data is preserved and the noise that is supposed to be isotropic is removed. Directional regularizers that are aware of dominant anisotropy directions in the image were proposed in [29, 32, 34, 37, 38]. While the model in [34] is limited to one dominant direction that is fixed a priori by the user, [29, 32] incorporate multiple directions of anisotropy in same spirit as the present paper, in the sense that they use an infimal convolution of multiple functionals, each with one dominant direction of anisotropy. But also in those works the bias toward a finite, user-defined number of directions remains. In [37, 38] the anisotropy is encoded in a space-dependent tensor that it is either estimated a priori or optimized in the algorithm by an alternating procedure. In contrast, our model is able to automatically select a distribution of directions and does not require a discretization of the parameter space; the set of directions is optimized together with the reconstruction through a provably convergent algorithm.

Outline of the main contributions:

- i) We introduce the *infinite infimal convolution functional* as a regularizer to solve ill-posed inverse problems defined in a Banach space X .
- ii) Under mild assumptions that are suitable for sparse regularization we prove well-posedness of the regularized inverse problem by constructing a convex lifting to the space of measures with values in X .
- iii) We use the convex lifting to gain insights into the structure of sparse solutions by proving a representer theorem in case of finite-dimensional data.
- iv) We propose a generalized conditional gradient method for solving the regularized problem, relying on the sparse structure of the iterate, and prove a sublinear rate of convergence.
- v) We demonstrate the performance of our model and the generalized conditional gradient method on several denoising problems using the infimal convolution of L^2 -norms of fractional-Laplacian-type operators.

2 The infinite infimal convolution functional

In this section we define the regularizer studied in this paper and state the corresponding regularized inverse problem.

2.1 Infinite infimal convolution

Let X be a separable Banach space that admits a predual X_* . Let Y be a separable Hilbert space and $A : X \rightarrow Y$ be a weak*-weak continuous operator (that is, A is the adjoint of some linear bounded operator $A_* : Y \rightarrow X_*$). Let $d \in \mathbb{N}$ and $S \subset \mathbb{R}^d$ be a non-empty compact set. We consider a proper Borel-measurable functional $J : X \times S \rightarrow [0, +\infty]$ satisfying the following assumptions:

(H1) $J(\cdot, s)$ is proper, convex, and positively one-homogeneous for every $s \in S$. Note that by positively one-homogeneous we mean that $J(\lambda v, s) = \lambda J(v, s)$ for every $\lambda \geq 0$, $v \in X$ and $s \in S$.

(H2) J is sequentially lower semicontinuous in $X \times S$, where X is endowed with the weak* topology.

(H3) There exists a constant $C > 0$ such that

$$\|v\|_X \leq C J(v, s) \quad (2.1)$$

for all $v \in X$ and for every $s \in S$.

Remark 2.1. Note that since X is separable, its predual X_* is separable as well. Moreover, thanks to Theorem A.4, X has the Radon–Nikodým-property, see Definition A.3.

Denote by $\mathcal{M}_+(S)$ the space of finite positive Radon measures on S . Given $\sigma \in \mathcal{M}_+(S)$ we define $F_\sigma : L^1((S, \sigma); X) \rightarrow [0, \infty]$ as follows

$$F_\sigma(u) := \int_S J(u(s), s) d\sigma(s), \quad (2.2)$$

where $L^1((S, \sigma); X)$ is the standard Bochner space of L^1 -functions with values in the Banach space X , cf. Section A.1. Note that the map $s \mapsto J(u(s), s)$ is σ -measurable for every $\sigma \in \mathcal{M}_+(S)$ and every $u \in L^1((S, \sigma); X)$, due to the measurability of J and the fact that σ is a Radon measure.

We are now ready to introduce the infinite infimal convolution functional.

Definition 2.2 (Infinite infimal convolution). The functional $R : X \times \mathcal{M}_+(S) \rightarrow [0, \infty]$ given by

$$R(v, \sigma) := \inf \left\{ F_\sigma(u) : u \in L^1((S, \sigma); X) \text{ with } \int_S u(s) d\sigma(s) = v \right\} \quad (2.3)$$

is called the infinite infimal convolution of the functionals $J(\cdot, s)$ over $s \in S$. The constraint $\int_S u(s) d\sigma(s) = v$ is defined as a Bochner integral.

In what follows we define for every $\sigma \in \mathcal{M}_+(S)$ an operator $\mathcal{I}_\sigma : L^1((S, \sigma); X) \rightarrow X$ as the integration of $u \in L^1((S, \sigma); X)$ against σ

$$\mathcal{I}_\sigma u := \int_S u(s) d\sigma(s).$$

With this notation we can rewrite the infinite infimal convolution functional equivalently as

$$R(v, \sigma) = \inf_{u \in L^1((S, \sigma); X)} \int_S J(u(s), s) d\sigma(s) + \mathbb{1}_{\{v\}}(\mathcal{I}_\sigma u), \quad (2.4)$$

where we denote by $\mathbb{1}_A : X \rightarrow [0, \infty]$ the indicator function of the Borel set $A \subset X$.

Remark 2.3. One can see that if we choose $\sigma = \sum_{i=1}^N \delta_{s_i}$ and $S = \{s_1, \dots, s_N\}$, where $N \in \mathbb{N}$ is fixed, the regularizer $R(v, \sigma)$ becomes the infimal convolution of the N functionals $v \mapsto J(v, s_i)$ for $i = 1, \dots, N$.

2.2 Tikhonov regularized inverse problems

For a given noisy measurement $f \in Y$ we are interested in inverse problems of the form

$$Av = f \tag{2.5}$$

where $A : X \rightarrow Y$ is a linear continuous operator.

In this paper we propose to solve (2.5) by setting up a Tikhonov problem regularized with the infinite infimal convolution functional. First, for a fixed $\sigma \in \mathcal{M}_+(S)$ we define the variational problem

$$\inf_{v \in X} \frac{1}{2} \|Av - f\|_Y^2 + \alpha R(v, \sigma), \tag{2.6}$$

where $\alpha > 0$ is a positive parameter. Since we are interested in optimizing (2.6) also over $\sigma \in \mathcal{M}_+(S)$, in order to find the best combination of functionals $J(\cdot, s)$ for $s \in S$ that optimally regularizes the reconstruction of a given measurement f , we additionally minimize (2.6) with respect to $\sigma \in \mathcal{M}_+(S)$ obtaining the following Tikhonov problem

$$\inf_{\substack{v \in X, \\ \sigma \in \mathcal{M}_+(S)}} \frac{1}{2} \|Av - f\|_Y^2 + \alpha R(v, \sigma). \tag{2.7}$$

Note that, using the definition of $R(v, \sigma)$, we can rewrite (2.7) as

$$\inf_{\substack{\sigma \in \mathcal{M}_+(S), \\ u \in L^1((S, \sigma); X)}} \frac{1}{2} \|A\mathcal{I}_\sigma u - f\|_Y^2 + \alpha \int_S J(u(s), s) d\sigma(s). \tag{2.8}$$

Clearly, (2.7) and (2.8) have the same infimum. Moreover, it is easy to check that if (u, σ) is a minimizer for (2.8), then $(\mathcal{I}_\sigma u, \sigma)$ is a minimizer for (2.7). Conversely, if (v, σ) is a minimizer for (2.7) and the infimum in (2.4) is achieved at $u \in L^1((S, \sigma); X)$, then (u, σ) is a minimizer for (2.8). It is our goal in the next section to show that these minima are indeed attained.

The lack of a predual of the space $L^1((S, \mu); X)$ and the relatively weak coercivity assumption (H3) do not allow one to show existence of minimizers for the infinite infimal convolution $R(v, \sigma)$ as in [41]. A possible solution is to assume that $\|v\|_X^p \leq C J(v, s)$ for every $s \in S$, $v \in X$ and $p > 1$. However, such additional assumption imposes a severe restriction on the class of regularizers one can consider, excluding, for example, sparsity-promoting regularizers. Our approach relies on the positive one-homogeneity of the functions $v \mapsto J(v, s)$ for each $s \in S$ (Assumption (H1)) and a lifting procedure to a space of X -valued measures over S . This yields a convex relaxation of (2.8) for which we can prove existence of minimizers. We also show that this implies existence of minimizers in (2.7) and (2.8).

3 Convex lifting

Since $R(v, \sigma)$ is not jointly convex in v, σ , the problem (2.7) is not convex. In this section we consider its convexification by lifting it to the space of X -valued Radon measures $\mathcal{M}(S; X)$. We collect necessary background on vector-valued measures in Appendix A.

3.1 Convex lifting of the regularizer

We first introduce a convex lifting of the map $(u, \sigma) \mapsto F_\sigma(u)$ defined in (2.2) to $\mathcal{M}(S; X)$. We define $\hat{F} : \mathcal{M}(S; X) \rightarrow [0, +\infty]$ as follows

$$\hat{F}(\mu) = \int_S J\left(\frac{\mu}{|\mu|}(s), s\right) d|\mu|(s), \tag{3.1}$$

where $|\mu| \in \mathcal{M}_+(S)$ is the variation measure of μ as defined in (A.3). Note that Remark 2.1 ensures the validity of the Radon–Nikodým property for X . Therefore, for every $\mu \in \mathcal{M}(S; X)$ the density $\frac{\mu}{|\mu|}$ exists and belongs to $L^1((S, \frac{\mu}{|\mu|}); X)$, cf. Definition A.3.

Proposition 3.1. *The functional $\hat{F}: \mathcal{M}(S; X) \rightarrow [0, +\infty]$ is convex, positively one-homogeneous and coercive, i.e., there exists $C > 0$ such that*

$$\hat{F}(\mu) \geq C\|\mu\|_{\mathcal{M}}, \quad (3.2)$$

for all $\mu \in \mathcal{M}(S; X)$. Moreover, given $\mu, \nu \in \mathcal{M}(S; X)$ such that $|\mu|$ and $|\nu|$ are mutually singular we have that

$$\hat{F}(\mu + \nu) = \hat{F}(\mu) + \hat{F}(\nu). \quad (3.3)$$

Proof. We start with the second statement. Let $\mu, \nu \in \mathcal{M}(S; X)$ be such that $|\mu|$ and $|\nu|$ are mutually singular. By mutual singularity, there exists a Borel set $A \subset S$ such that $|\nu|(A) = 0$ and $|\mu|(S \setminus A) = 0$. We now show that $\frac{\mu + \nu}{|\mu + \nu|} = \frac{\mu}{|\mu|}$ holds $|\mu|$ -a.e. in A . For every $E \subset A$ Borel we observe that

$$\int_E \frac{\mu + \nu}{|\mu + \nu|}(s) d|\mu + \nu|(s) = \mu(E) + \nu(E) = \mu(E) = \int_E \frac{\mu}{|\mu|}(s) d|\mu|(s) \quad (3.4)$$

and

$$|\mu + \nu|(E) = \sup \left\{ \sum_{i=1}^n \|\mu(A_i)\|_X : n \in \mathbb{N}, (A_i)_{i=1}^n \text{ partition of } E \right\} = |\mu|(E). \quad (3.5)$$

From (3.4) and (3.5) and Corollary 5 in [25] we conclude that $\frac{\mu + \nu}{|\mu + \nu|} = \frac{\mu}{|\mu|}$ for $|\mu|$ -a.e. in A . Similarly, it can be shown that $\frac{\mu + \nu}{|\mu + \nu|} = \frac{\nu}{|\nu|}$ for $|\nu|$ -a.e. in $S \setminus A$. So,

$$\begin{aligned} \hat{F}(\mu + \nu) &= \int_S J \left(\frac{\mu + \nu}{|\mu + \nu|}(s), s \right) d|\mu + \nu|(s) \\ &= \int_A J \left(\frac{\mu + \nu}{|\mu + \nu|}(s), s \right) d|\mu + \nu|(s) + \int_{S \setminus A} J \left(\frac{\mu + \nu}{|\mu + \nu|}(s), s \right) d|\mu + \nu|(s) \\ &= \int_A J \left(\frac{\mu}{|\mu|}(s), s \right) d|\mu|(s) + \int_{S \setminus A} J \left(\frac{\nu}{|\nu|}(s), s \right) d|\nu|(s) = \hat{F}(\mu) + \hat{F}(\nu), \end{aligned}$$

proving (3.3).

Now, it is clear that \hat{F} is positively one-homogeneous. Indeed, by the definition of \hat{F} one immediately sees that $\hat{F}(0) = 0$. Moreover, for every $\lambda > 0$, Proposition A.2 yields that $\frac{\lambda\mu}{|\lambda\mu|}(s) = \frac{\mu}{|\mu|}(s)$ for $|\mu|$ -a.e. $s \in S$ and hence,

$$\hat{F}(\lambda\mu) = \int_S J \left(\frac{\mu}{|\mu|}(s), s \right) d(|\lambda\mu|)(s) = \lambda\hat{F}(\mu).$$

Since \hat{F} is positively one-homogeneous, convexity is equivalent to subadditivity, i.e., it is enough to show that $\hat{F}(\mu + \nu) \leq \hat{F}(\mu) + \hat{F}(\nu)$ for all $\mu, \nu \in \mathcal{M}(S; X)$. Let $\mu, \nu \in \mathcal{M}(S; X)$ be arbitrary. First note that

$$\frac{\mu + \nu}{|\mu + \nu|}(s) = \frac{\mu + \nu}{|\mu + \nu|}(s) \frac{|\mu + \nu|}{|\mu + \nu|}(s) \quad \text{for } (|\mu| + |\nu|)\text{-a.e. } s \in S, \quad (3.6)$$

where all densities are well-defined since X has the Radon–Nikodým property, cf. Remark 2.1. Indeed, since $|\mu + \nu| = \frac{|\mu + \nu|}{|\mu + \nu|}(|\mu| + |\nu|)$ by Radon–Nikodým property, we have that

$$\begin{aligned} \int_E \frac{\mu + \nu}{|\mu + \nu|}(s) d(|\mu| + |\nu|)(s) &= (\mu + \nu)(E) = \int_E \frac{\mu + \nu}{|\mu + \nu|}(s) d(|\mu + \nu|)(s) \\ &= \int_E \frac{\mu + \nu}{|\mu + \nu|}(s) \frac{|\mu + \nu|}{|\mu + \nu|}(s) d(|\mu| + |\nu|)(s) \end{aligned}$$

for every $E \subset S$ Borel, showing (3.6) thanks to [25, Corollary 5]. Similarly, it also holds that

$$\frac{\mu}{|\mu| + |\nu|}(s) = \frac{\mu}{|\mu|}(s) \frac{|\mu|}{|\mu| + |\nu|}(s) \quad \text{for } (|\mu| + |\nu|) - a.e. \ s \in S, \quad (3.7)$$

$$\frac{\nu}{|\mu| + |\nu|}(s) = \frac{\nu}{|\nu|}(s) \frac{|\nu|}{|\mu| + |\nu|}(s) \quad \text{for } (|\mu| + |\nu|) - a.e. \ s \in S, \quad (3.8)$$

$$\frac{\mu + \nu}{|\mu| + |\nu|}(s) = \frac{\mu}{|\mu| + |\nu|}(s) + \frac{\nu}{|\mu| + |\nu|}(s) \quad \text{for } (|\mu| + |\nu|) - a.e. \ s \in S. \quad (3.9)$$

Therefore,

$$\begin{aligned} \hat{F}(\mu + \nu) &= \int_S J\left(\frac{\mu + \nu}{|\mu + \nu|}(s), s\right) d|\mu + \nu|(s) = \int_S J\left(\frac{\mu + \nu}{|\mu + \nu|}(s), s\right) \frac{|\mu + \nu|}{|\mu| + |\nu|}(s) d(|\mu| + |\nu|)(s) \\ &= \int_S J\left(\frac{\mu + \nu}{|\mu| + |\nu|}(s), s\right) d(|\mu| + |\nu|)(s) \\ &\leq \int_S J\left(\frac{\mu}{|\mu| + |\nu|}(s), s\right) d(|\mu| + |\nu|)(s) + \int_S J\left(\frac{\nu}{|\mu| + |\nu|}(s), s\right) d(|\mu| + |\nu|)(s) \\ &= \int_S J\left(\frac{\mu}{|\mu|}(s), s\right) \frac{|\mu|}{|\mu| + |\nu|}(s) d(|\mu| + |\nu|)(s) + \int_S J\left(\frac{\nu}{|\nu|}(s), s\right) \frac{|\nu|}{|\mu| + |\nu|}(s) d(|\mu| + |\nu|)(s) \\ &= \hat{F}(\mu) + \hat{F}(\nu), \end{aligned}$$

where in the second equality we applied Radon–Nikodým’s theorem and in the third equality we used (3.6) and the positive one-homogeneity of $J(\cdot, s)$, noting that $\frac{|\mu + \nu|}{|\mu| + |\nu|}$ is non-negative for $(|\mu| + |\nu|)$ -a.e. $s \in S$; additionally, in the first inequality we used the subadditivity of $J(\cdot, s)$ and (3.9), and in fourth equality we used (3.7), (3.8) and the positive one-homogeneity of $J(\cdot, s)$, noting that both $\frac{|\mu|}{|\mu| + |\nu|}$ and $\frac{|\nu|}{|\mu| + |\nu|}$ are non-negative for $(|\mu| + |\nu|)$ -a.e. $s \in S$; finally the last equality follows again from Radon–Nikodým’s theorem.

To show coercivity (3.2), we use Assumption (H3) and Proposition A.2, and obtain

$$\hat{F}(\mu) = \int_S J\left(\frac{\mu}{|\mu|}(s), s\right) d|\mu|(s) \geq \int_S \frac{1}{C} \left\| \frac{\mu}{|\mu|}(s) \right\|_X d|\mu|(s) = \frac{1}{C} \|\mu\|_{\mathcal{M}},$$

where C is the constant from Assumption (H3). \square

Next we establish weak* lower semicontinuity of the functional $\hat{F}: \mathcal{M}(S; X) \rightarrow [0, +\infty]$. We remind the reader that the weak* topology on the space $\mathcal{M}(S; X)$ is understood in the sense of duality with $C(S; X_*)$, see Theorem A.6. The proof uses the same techniques as in the proof of the classical Reshetnyak semicontinuity theorem, see [40] and [1, Theorem 2.38], adapted to measures with values in Banach spaces. The same result is proven in [21, Theorem 8.2.2] for a separable and reflexive space X . For completeness, we provide a complete proof.

Theorem 3.2. *The functional \hat{F} is weakly-* sequentially lower semicontinuous in $\mathcal{M}(S; X)$.*

Proof. Let $\mu^n \xrightarrow{*} \mu$ in $\mathcal{M}(S; X)$, i.e., for any $\varphi \in C(S; X_*)$,

$$\langle \mu^n, \varphi \rangle \rightarrow \langle \mu, \varphi \rangle.$$

By the Radon–Nikodým theorem, there exists a sequence $\{u^n\}_n$ of Borel measurable functions such that $u^n \in L^1((S, |\mu^n|); X)$ and an element $u \in L^1((S, |\mu|); X)$ such that $\mu^n = u^n |\mu^n|$ for all n and $\mu = u |\mu|$. By Proposition A.2 we can always assume that $\|u^n(s)\|_X = 1$ for $|\mu^n|$ -almost every $s \in S$. Since weakly-* convergent sequences are bounded, we have $\sup_n \|\mu^n\|_{\mathcal{M}} < \infty$.

Denote by $B_X := \{u \in X : \|u\|_X \leq 1\}$ the unit ball in X . By the Banach–Alaoglu theorem, B_X is weakly-* compact and thanks to the separability of X , the weak* topology is metrizable. Since S

is compact, the metric space $B_X \times S$ (where B_X is endowed with the metric that metrizes its weak* topology) is compact and therefore separable. Let the sequence $\{\nu^n\}_n$ in $\mathcal{M}(B_X \times S)$ be defined as

$$\int_{B_X \times S} \varphi(x, s) d\nu^n(x, s) = \int_S \varphi(u^n(s), s) d|\mu^n|(s) \quad \forall \varphi \in C(B_X \times S). \quad (3.10)$$

Since, using (3.10), we have that $\sup_n \|\nu^n\|_{\mathcal{M}} = \sup_n \|\mu^n\|_{\mathcal{M}} < \infty$, the sequence $\{\nu^n\}_n$ is bounded and therefore contains a weakly-* converging subsequence (which we do not relabel)

$$\nu^n \xrightarrow{*} \nu \in \mathcal{M}(B_X \times S).$$

Let $\pi : X \times S \rightarrow S$ be the projection on S . Then we have $\pi_{\#}\nu^n = |\mu^n|$ and $|\mu^n| \xrightarrow{*} \pi_{\#}\nu$ by the continuity of the push-forward with respect to weak* convergence. Moreover, defining $\lambda := \pi_{\#}\nu$ and applying [1, Proposition 1.62], we get that $\lambda \geq |\mu|$. Applying the disintegration theorem for measures defined on metric spaces (see for example [11, Theorem A.4]), there exists a Borel family of measures $\{\nu_s\}_{s \in S}$ in $\mathcal{M}(B_X)$ such that for every $f \in L^1(B_X \times S, \nu)$

$$\int_{B_X \times S} f(x, s) d\nu(x, s) = \int_S \int_{B_X} f(x, s) d\nu_s(x) d\lambda(s). \quad (3.11)$$

We now analyze the barycenter of ν_s in B_X for every $s \in S$, that is

$$\text{bar}(\nu_s) := \int_{B_X} x d\nu_s(x),$$

where the previous integral is well-defined as a Bochner integral since the identity map is continuous on the separable metric space B_X and hence, is in $L^1(B_X, \nu_s)$ for every $s \in S$.

Let us define $f : (x, s) \mapsto \langle \psi(s), x \rangle \in C(B_X \times S)$, where $\psi \in C(S; X_*)$. Using (3.11) and standard properties of the Bochner integral we get the following equality for the sequences $\nu^n \xrightarrow{*} \nu$ and $\mu^n \xrightarrow{*} \mu$:

$$\begin{aligned} \int_S \langle \psi(s), \int_{B_X} x d\nu_s(x) \rangle d\lambda(s) &= \int_S \int_{B_X} \langle \psi(s), x \rangle d\nu_s(x) d\lambda(s) = \int_{B_X \times S} \langle \psi(s), x \rangle d\nu(x, s) \\ &= \lim_{n \rightarrow +\infty} \int_{B_X \times S} \langle \psi(s), x \rangle d\nu^n(x, s) = \lim_{n \rightarrow +\infty} \int_S \langle \psi(s), u^n(s) \rangle d|\mu^n|(s) \\ &= \lim_{n \rightarrow +\infty} \int_S \psi(s) d\mu^n(s) = \int_S \psi(s) d\mu(s) = \int_S \psi(s) u(s) \frac{|\mu|}{\lambda}(s) d\lambda(s). \end{aligned}$$

In particular, it holds that

$$\text{bar}(\nu_s) = u(s) \frac{|\mu|}{\lambda}(s) \quad \text{for } \lambda - a.e. s \in S. \quad (3.12)$$

Therefore, by [11, Proposition A.3] (see also [2, Section 5.1.7]) and the lower semicontinuity of J on $B_X \times S$ (see Assumption (H2)) we have

$$\begin{aligned} \liminf_{n \rightarrow +\infty} \int_S J(u^n(s), s) d|\mu^n|(s) &= \liminf_{n \rightarrow +\infty} \int_{B_X \times S} J(x, s) d\nu^n(x, s) \geq \int_{B_X \times S} J(x, s) d\nu(x, s) \\ &= \int_S \int_{B_X} J(x, s) d\nu_s(x) d\lambda(s) \geq \int_S J\left(u(s) \frac{|\mu|}{\lambda}(s), s\right) d\lambda(s) = \int_S J(u(s), s) d|\mu|(s), \end{aligned}$$

where we used (3.11), (3.12), Jensen's inequality (see Assumption (H1)). The proof is complete. \square

We conclude the section showing that $\hat{F}(\mu)$ coincides with $\int_S J(u(s), s) d\sigma(s)$ when computed on measures $\mu \in \mathcal{M}(S; X)$ of the form $\mu(E) = \int_E u(s) d\sigma(s)$ for $E \subset S$ Borel, where $\sigma \in \mathcal{M}_+(S)$ and $u \in L^1((S, \sigma); X)$.

Lemma 3.3. *Let $\sigma \in \mathcal{M}_+(S)$ and $u \in L^1((S, \sigma); X)$. Then, for $\mu \in \mathcal{M}(S; X)$ defined as $\mu(E) = \int_E u(s) d\sigma(s)$, $E \subset S$ Borel, it holds that*

$$\hat{F}(\mu) = \int_S J(u(s), s) d\sigma(s). \quad (3.13)$$

Proof. Using Proposition A.2, the fact that $\|u\|_X \neq 0$ $|\mu|$ -a.e., and the positive one-homogeneity of $J(\cdot, s)$ for every $s \in S$ (see Assumption (H1)), we get

$$\hat{F}(\mu) = \int_S J\left(\frac{u(s)}{\|u(s)\|_X}, s\right) \|u(s)\|_X d\sigma(s) = \int_S J(u(s), s) d\sigma(s)$$

as we wanted to prove. \square

3.2 Convex lifting of the variational problem

We are now ready to lift the problem (2.7) (and therefore also (2.8)) to the space of X -valued measures $\mathcal{M}(S; X)$. Consider

$$\inf_{\mu \in \mathcal{M}(S; X)} \frac{1}{2} \|A\mu(S) - f\|_Y^2 + \alpha \int_S J\left(\frac{\mu}{|\mu|}(s), s\right) d|\mu|(s). \quad (3.14)$$

The following result shows the equivalence between the Tikhonov regularized inverse problem (2.8) and the lifted problem (3.14).

Proposition 3.4. *Problems (2.8) and (3.14) have the same infimum. Moreover, if the pair $(\bar{u}, \bar{\sigma})$ minimizes (2.8), then $\bar{\mu}$ defined as*

$$\bar{\mu}(E) = \int_E \bar{u}(s) d\bar{\sigma}(s) \quad (3.15)$$

is a minimizer for (3.14). Conversely, if $\bar{\mu}$ is a minimizer for (3.14), then the pair $\left(\frac{\bar{\mu}}{|\bar{\mu}|}, |\bar{\mu}|\right)$ is a minimizer for (2.8) and $(\bar{\mu}(S), |\bar{\mu}|)$ is a minimizer for (2.7).

Proof. Denote by $(\bar{u}, \bar{\sigma})$ a pair minimizing (2.8) and define $\bar{\mu} \in \mathcal{M}(S; X)$ as in (3.15). Given $\mu \in \mathcal{M}(S; X)$, since X satisfies the Radon–Nikodým property, we have

$$\mu(E) = \int_E \frac{\mu}{|\mu|}(s) d|\mu|(s), \quad (3.16)$$

so that $\mathcal{I}_{|\mu|} \frac{\mu}{|\mu|} = \mu(S)$. Therefore, using Lemma 3.3 and optimality we have

$$\begin{aligned} & \frac{1}{2} \|A\mu(S) - f\|_Y^2 + \alpha \int_S J\left(\frac{\mu}{|\mu|}(s), s\right) d|\mu|(s) \\ &= \frac{1}{2} \left\| A\mathcal{I}_{|\mu|} \frac{\mu}{|\mu|} - f \right\|_Y^2 + \alpha \int_S J\left(\frac{\mu}{|\mu|}(s), s\right) d|\mu|(s) \\ &\geq \frac{1}{2} \|A\mathcal{I}_{\bar{\sigma}} \bar{u} - f\|_Y^2 + \alpha \int_S J(\bar{u}(s), s) d\bar{\sigma}(s) \\ &= \frac{1}{2} \|A\bar{\mu}(S) - f\|_Y^2 + \alpha \int_S J\left(\frac{\bar{\mu}}{|\bar{\mu}|}(s), s\right) d|\bar{\mu}|(s), \end{aligned}$$

implying that $\bar{\mu} \in \mathcal{M}(S; X)$ as in (3.15) is a minimizer of (3.14).

Conversely, let $\bar{\mu}$ be a minimizer for (3.14). Consider $\sigma \in \mathcal{M}_+(S)$ and $u \in L^1((S, \sigma); X)$, and define the measure $\mu \in \mathcal{M}(S; X)$ as $\mu(E) := \int_E u(s) d\sigma(s)$ for every Borel subset $E \subset S$. Setting $(\bar{u}, \bar{\sigma}) = \left(\frac{\bar{\mu}}{|\bar{\mu}|}, |\bar{\mu}|\right)$, it holds true that

$$\bar{\mu}(S) = \int_S \frac{\bar{\mu}}{|\bar{\mu}|}(s) d|\bar{\mu}|(s) = \int_S \bar{u}(s) d\bar{\sigma}(s)$$

and we have, again using Lemma 3.3 and optimality,

$$\begin{aligned}
\frac{1}{2}\|A\mathcal{I}_\sigma u - f\|_Y^2 + \alpha \int_S J(u(s), s) d\sigma(s) &= \frac{1}{2}\|A\mu(S) - f\|_Y^2 + \alpha \int_S J\left(\frac{\mu}{|\mu|}(s), s\right) d|\mu|(s) \\
&\geq \frac{1}{2}\|A\bar{\mu}(S) - f\|_Y^2 + \alpha \int_S J\left(\frac{\bar{\mu}}{|\bar{\mu}|}(s), s\right) d|\bar{\mu}|(s) \\
&= \frac{1}{2}\|A\mathcal{I}_{\bar{\sigma}}\bar{u} - f\|_Y^2 + \alpha \int_S J(\bar{u}(s), s) d\bar{\sigma}(s),
\end{aligned}$$

implying that $\left(\frac{\bar{\mu}}{|\bar{\mu}|}, |\bar{\mu}|\right)$ is a minimizer for (2.8) and therefore the pair $(\mathcal{I}_{|\bar{\mu}|}\frac{\bar{\mu}}{|\bar{\mu}|}, |\bar{\mu}|) = (\bar{\mu}(S), |\bar{\mu}|)$ is a minimizer for (2.7).

Finally, we show that (2.8) and (3.14) have the same infimum. Given any $\mu \in \mathcal{M}(S; X)$ define $\tilde{\sigma} := |\mu| \in \mathcal{M}_+(S)$ and $\tilde{u} := \frac{\mu}{|\mu|} \in L^1((S, \tilde{\sigma}); X)$. Then, with similar computations to the first part of the proof, we have that

$$\begin{aligned}
\inf_{\substack{\sigma \in \mathcal{M}_+(S), \\ u \in L^1((S, \sigma); X)}} \frac{1}{2}\|A\mathcal{I}_\sigma u - f\|_Y^2 + \alpha \int_S J(u(s), s) d\sigma(s) &\leq \frac{1}{2}\|A\mathcal{I}_{\tilde{\sigma}}\tilde{u} - f\|_Y^2 + \alpha \int_S J(\tilde{u}(s), s) d\tilde{\sigma}(s) \\
&= \frac{1}{2}\|A\mu(S) - f\|_Y^2 + \alpha \int_S J\left(\frac{\mu}{|\mu|}(s), s\right) d|\mu|(s),
\end{aligned}$$

implying that the infimum of (2.8) is less or equal than the infimum of (3.14). Similarly, given any $\sigma \in \mathcal{M}_+(S)$ and $u \in L^1((S, \sigma); X)$, and defining $\tilde{\mu} := u\sigma \in \mathcal{M}(S; X)$ there holds

$$\begin{aligned}
\inf_{\mu \in \mathcal{M}(S; X)} \frac{1}{2}\|A\mu(S) - f\|_Y^2 + \alpha \int_S J\left(\frac{\mu}{|\mu|}, s\right) d|\mu|(s) &\leq \frac{1}{2}\|A\tilde{\mu}(S) - f\|_Y^2 + \alpha \int_S J\left(\frac{\tilde{\mu}}{|\tilde{\mu}|}(s), s\right) d|\tilde{\mu}|(s) \\
&= \frac{1}{2}\|A\mathcal{I}_\sigma u - f\|_Y^2 + \alpha \int_S J(u(s), s) d\sigma(s),
\end{aligned}$$

implying that the infimum of (3.14) is less or equal than the infimum of (2.8) and thus concluding the proof. \square

Remark 3.5. We recall that Problems (2.7) and (2.8) are equivalent as discussed in Section 2.2. Therefore, Proposition 3.4 implies that (2.7), (2.8) and (3.14) have the same infimum and, given a minimizer of one of these problems, one can construct explicitly a minimizer of the others. In the next section we make sure that such minimizers actually exist.

3.3 Existence of minimizers

We are now ready to show that (2.7) and (2.8) admit a minimizer. First we show that the lifted problem (3.14) admits a solution and then thanks to Proposition 3.4 we immediately infer that (2.8) and (2.7) admit minimizers as well.

Proposition 3.6. *There exists $\mu \in \mathcal{M}(S; X)$ that minimizes the lifted problem (3.14).*

Proof. Consider a minimizing sequence $\{\mu^n\}_n$ in $\mathcal{M}(S; X)$ for (3.14). Then, thanks to Proposition 3.1, the sequence $\{\mu^n\}_n$ is uniformly bounded in total variation. Hence, by the Banach–Alaoglu theorem and Theorem A.6, there exists $\mu \in \mathcal{M}(S; X)$ such that, up to a subsequence, $\mu^n \xrightarrow{*} \mu$. Thanks to the weak*-to-weak continuity of A , applying Remark A.7 and the weak* lower semicontinuity of \hat{F} provided by Theorem 3.2, we conclude that μ is a minimizer of (3.14). \square

Corollary 3.7. *There exists a pair (u, σ) with $\sigma \in \mathcal{M}_+(S)$ and $u \in L^1((S, \sigma); X)$ that minimizes (2.8). Consequently, the pair $(\mathcal{I}_\sigma u, \sigma) \in X \times \mathcal{M}_+(S)$ minimizes (2.7).*

Proof. The proof follows directly from Proposition 3.4 and Proposition 3.6. \square

3.4 Extremal points of the lifted regularizer

We conclude this section by characterizing the set of extremal points of the unit ball of $\alpha\hat{F}$, that is

$$B := \{\mu \in \mathcal{M}(S; X) : \alpha\hat{F}(\mu) \leq 1\}, \quad (3.17)$$

in terms of the extremal points of

$$B_s := \{v \in X : \alpha J(v, s) \leq 1\} \quad (3.18)$$

for $s \in S$ where we do not highlight the dependence on the parameter α in the definition of the unit balls. Note that the choice to consider unit ball of $\alpha\hat{F}$, instead of \hat{F} , is made purely for notational convenience, since the generalized conditional gradient method introduced in [13] is formulated for composite minimization problems without explicit regularization parameters. This result will be fundamental in Section 4 to prove a representer theorem for inverse problems regularized with the infinite infimal convolution functional and in Section 5 to set up a generalized conditional gradient method for its numerical solution. Note that while this result is specific for infinite infimal convolution functionals, results concerning the extremal points of unit balls of classical infimal convolution functionals can be found in [20, 33]

Theorem 3.8. *The extremal points of B can be characterized as follows:*

- If $0 \in \text{Ext}(B_r)$ for each $r \in S$, then $\text{Ext}(B) = \{v\delta_r : r \in S, v \in \text{Ext}(B_r)\}$.
- If there exists $r \in S$ such that $0 \notin \text{Ext}(B_r)$, then $\text{Ext}(B) = \{v\delta_r : r \in S, v \in \text{Ext}(B_r) \setminus \{0\}\}$.

Proof. Note that it is enough to prove the statement for $\alpha = 1$, since, due the positive one-homogeneity of $J(\cdot, s)$ for every $s \in S$ and of \hat{F} , the general case follows from a scaling argument. Given $r \in S$ and $v \in B_r$, notice that $|v\delta_r|(E) = \int_E \|v(s)\|_X d\delta_r(s) = \|v\|_X \delta_r(E)$. Therefore, using the positive one-homogeneity of J , we get

$$\hat{F}(v\delta_r) = \int_S J(v, s) d\delta_r(s) = J(v, r) \leq 1, \quad (3.19)$$

since $v \in B_r$. Thus, $v\delta_r \in B$.

First, we show that $\text{Ext}(B) \subset \{v\delta_r : r \in S, v \in \text{Ext}(B_r)\}$ and in case there exists $\hat{r} \in S$ such that $0 \notin \text{Ext}(B_{\hat{r}})$ it also holds that $\text{Ext}(B) \subset \{v\delta_r : r \in S, v \in \text{Ext}(B_r) \setminus \{0\}\}$. Note that, given $\mu \in \text{Ext}(B)$, either $\hat{F}(\mu) = 0$ or $\hat{F}(\mu) = 1$, since \hat{F} is positively one-homogeneous. Suppose that $\hat{F}(\mu) = 0$, so that $\mu = 0$ by Proposition 3.1. If $0 \in \text{Ext}(B_r)$ for each $r \in S$, then we immediately have that $0 \in \{v\delta_r : r \in S, v \in \text{Ext}(B_r)\}$ as we wanted to prove. Suppose instead that there exists $\hat{r} \in S$ such that $0 \notin \text{Ext}(B_{\hat{r}})$. Then,

$$0 = \lambda v_1 + (1 - \lambda)v_2 \quad (3.20)$$

for $v_1, v_2 \in B_{\hat{r}}$, $v_1 \neq v_2$ and $0 < \lambda < 1$. Therefore the convex decomposition $0 = \lambda v_1 \delta_{\hat{r}} + (1 - \lambda)v_2 \delta_{\hat{r}}$ shows that 0 is not an extremal point of the ball \hat{F} , since $\hat{F}(v_i \delta_{\hat{r}}) = J(v_i, \hat{r}) \leq 1$.

Suppose now that $\hat{F}(\mu) = 1$. Note that, since $\hat{F}(\mu) = 1$ implies that $\mu \neq 0$, it is enough to show that $\mu \in \{v\delta_r : r \in S, v \in \text{Ext}(B_r)\}$. We start by showing that μ is supported on a singleton. If this does not hold, then there exists a Borel set $A \subset S$ such that $0 < \mu(A) < \mu(S)$. Define the constants

$$C(A) := \int_A J\left(\frac{\mu}{|\mu|}(s), s\right) d|\mu|(s) \quad \text{and} \quad C(S \setminus A) := \int_{S \setminus A} J\left(\frac{\mu}{|\mu|}(s), s\right) d|\mu|(s).$$

Since $J(v, s) \geq C\|v\|_X$ for each $v \in X$, $s \in S$ by (H3), there exists a constant $c > 0$ such that $C(A) \geq c|\mu|(A) > 0$ and $C(S \setminus A) \geq c|\mu|(S \setminus A) > 0$. Moreover, since $\hat{F}(\mu) = 1$, it holds that $C(A) + C(S \setminus A) = 1$ and in particular, they are both finite. Defining then the measures

$$\mu_1 := \frac{\mu \llcorner A}{C(A)} \quad \text{and} \quad \mu_2 := \frac{\mu \llcorner (S \setminus A)}{C(S \setminus A)},$$

we notice that

$$|\mu_1| = \frac{|\mu| \llcorner A}{C(A)}, \quad \frac{\mu_1}{|\mu_1|} = \frac{\mu}{|\mu|} \quad |\mu_1| - a.e., \quad |\mu_2| = \frac{|\mu| \llcorner (S \setminus A)}{C(S \setminus A)}, \quad \frac{\mu_2}{|\mu_2|} = \frac{\mu}{|\mu|} \quad |\mu_2| - a.e..$$

In particular, we have that $\hat{F}(\mu_1) = \hat{F}(\mu_2) = 1$ and thus

$$\mu = C(A)\mu_1 + C(S \setminus A)\mu_2$$

gives a non-trivial decomposition, which is a contradiction to $\mu \in \text{Ext}(B)$. Therefore, μ is supported on a singleton and thus, $\mu = v\delta_r$ for some $v \in X$ and $r \in S$. We now show that $v \in \text{Ext}(B_r)$. If this does not hold, then there exist $0 < \lambda < 1$ and $v_1, v_2 \in B_r$ such that $v_1 \neq v_2$ and

$$v = \lambda v_1 + (1 - \lambda)v_2. \quad (3.21)$$

Multiplying the previous equation by δ_r we obtain that $\mu = \lambda v_1 \delta_r + (1 - \lambda)v_2 \delta_r$ and, since $v_1, v_2 \in B_r$ and $\hat{F}(v_i \delta_r) = J(v_i, r) \leq 1$ for $i = 1, 2$, we conclude that μ is not an extremal point of B , which is a contradiction.

We now show that $\text{Ext}(B) \supset \{v\delta_r : r \in S, v \in \text{Ext}(B_r) \setminus \{0\}\}$. Consider $r \in S$, $v \in \text{Ext}(B_r) \setminus \{0\}$, $0 < \lambda < 1$ and $\mu_1, \mu_2 \in B$ such that

$$v\delta_r = \lambda\mu_1 + (1 - \lambda)\mu_2. \quad (3.22)$$

Note that, since $v \neq 0$ and $J(\cdot, r)$ is coercive (Assumption **(H3)**) and positively one-homogeneous, we have $J(v, r) = 1$, as otherwise $v \notin \text{Ext}(B_r)$. Therefore

$$\begin{aligned} 1 = J(v, r) &= \hat{F}(v\delta_r) = \hat{F}([\lambda\mu_1 + (1 - \lambda)\mu_2] \llcorner \{r\}) + \hat{F}([\lambda\mu_1 + (1 - \lambda)\mu_2] \llcorner (S \setminus \{r\})) \\ &\leq \lambda \hat{F}(\mu_1 \llcorner \{r\}) + (1 - \lambda) \hat{F}(\mu_2 \llcorner \{r\}) \leq 1, \end{aligned} \quad (3.23)$$

where in the third equality we apply **(3.3)** in Proposition **3.1**, in the first inequality we use that $\hat{F}(0) = 0$ as well as the convexity of \hat{F} , and in the last inequality we use that $\hat{F}(\mu_j \llcorner \{r\}) \leq \hat{F}(\mu_j)$ for $j = 1, 2$, that is true due to **(3.3)**, since $\hat{F} \geq 0$. Note now that $\mu_j \llcorner \{r\} = v_j \delta_r$ with $v_j \in B_r$ for $j = 1, 2$, and $\hat{F}(v_j \delta_r) = J(v_j, r) = 1$ for $j = 1, 2$, since otherwise we would contradict **(3.23)**. Moreover, thanks to the estimate $1 \leq \hat{F}(\mu_j \llcorner (S \setminus \{r\})) + \hat{F}(v_j \delta_r) = \hat{F}(\mu_j) \leq 1$, we deduce that $\hat{F}(\mu_j \llcorner (S \setminus \{r\})) = 0$ and thus $\mu_j \llcorner (S \setminus \{r\}) = 0$ for $j = 1, 2$, due to **(3.2)** in Proposition **3.1**. In particular, we obtain $\mu_j = v_j \delta_r$ for $j = 1, 2$. Finally, substituting $\mu_j = v_j \delta_r$ for $j = 1, 2$ in **(3.22)** we deduce that $v = \lambda v_1 + (1 - \lambda)v_2$, which implies that $v_1 = v_2 = v$, due to the extremality of v . We then conclude that $\mu_1 = \mu_2$, implying that $v\delta_r \in \text{Ext}(B)$.

We conclude the proof by showing that if $0 \in \text{Ext}(B_r)$ for every $r \in S$, then $0 \in \text{Ext}(B)$. If $0 = \lambda\mu_1 + (1 - \lambda)\mu_2$ with $0 < \lambda < 1$ and $\mu_1, \mu_2 \in B$, note that $\mu_j \llcorner \llcorner |\mu_1| + |\mu_2|$ for $j = 1, 2$. Therefore, applying the Radon–Nikodým property of X , cf. Definition **A.3**, we have that for $j = 1, 2$ it holds

$$\begin{aligned} 1 \geq \hat{F}(\mu_j) &= \int_S J\left(\frac{\mu_j}{|\mu_j|}(s), s\right) d|\mu_j|(s) = \int_S J\left(\frac{\mu_j}{|\mu_j|}(s), s\right) \frac{|\mu_j|}{|\mu_1| + |\mu_2|}(s) d(|\mu_1| + |\mu_2|)(s) \\ &= \int_S J\left(\frac{\mu_j}{|\mu_1| + |\mu_2|}(s), s\right) d(|\mu_1| + |\mu_2|)(s), \end{aligned}$$

where we have also used the positive one-homogeneity of $J(\cdot, s)$ and **(3.7)**, **(3.8)** with $\mu_1 = \mu$ and $\mu_2 = \nu$. In particular, $J\left(\frac{\mu_j}{|\mu_1| + |\mu_2|}(s), s\right) < +\infty$ for $(|\mu_1| + |\mu_2|)$ -a.e. $s \in S$ and $j = 1, 2$. Now, from the convex decomposition $0 = \lambda\mu_1 + (1 - \lambda)\mu_2$, by applying the Radon–Nikodým property of X we have that $0 = \left(\lambda \frac{\mu_1}{|\mu_1| + |\mu_2|} + (1 - \lambda) \frac{\mu_2}{|\mu_1| + |\mu_2|}\right) (|\mu_1| + |\mu_2|)$ and thus

$$0 = \lambda \frac{\mu_1}{|\mu_1| + |\mu_2|}(r) + (1 - \lambda) \frac{\mu_2}{|\mu_1| + |\mu_2|}(r) \quad \text{for } (|\mu_1| + |\mu_2|) - a.e. \ r \in S. \quad (3.24)$$

Define $c(r) = J\left(\frac{\mu_1}{|\mu_1|+|\mu_2|}(r), r\right) + J\left(\frac{\mu_2}{|\mu_1|+|\mu_2|}(r), r\right) < \infty$ for $(|\mu_1|+|\mu_2|)$ -a.e. $r \in S$. By Assumption (H3), it holds that $c(r) > 0$ for $(|\mu_1|+|\mu_2|)$ -a.e. $r \in S$ such that $\frac{\mu_1}{|\mu_1|+|\mu_2|}(r) \neq 0$ or $\frac{\mu_2}{|\mu_1|+|\mu_2|}(r) \neq 0$. Therefore, up to setting $c(r)$ equal to an arbitrary positive number on the points such that $\frac{\mu_j}{|\mu_1|+|\mu_2|}(r) = 0$ for $j = 1, 2$, we obtain from (3.24) that

$$0 = \lambda \frac{1}{c(r)} \frac{\mu_1}{|\mu_1|+|\mu_2|}(r) + (1-\lambda) \frac{1}{c(r)} \frac{\mu_2}{|\mu_1|+|\mu_2|}(r) \quad \text{for } (|\mu_1|+|\mu_2|) - \text{a.e. } r \in S. \quad (3.25)$$

Thanks to the positive one-homogeneity of $J(\cdot, r)$, we have that $\frac{1}{c(r)} \frac{\mu_j}{|\mu_1|+|\mu_2|}(r) \in B_r$ for $(|\mu_1|+|\mu_2|)$ -a.e. $r \in S$, $j = 1, 2$. Therefore, since $0 \in \text{Ext}(B_r)$ for every $r \in S$, it holds that $0 = \frac{\mu_1}{|\mu_1|+|\mu_2|}(r) = \frac{\mu_2}{|\mu_1|+|\mu_2|}(r)$ for $(|\mu_1|+|\mu_2|)$ -a.e. $r \in S$ and thus $\mu_j = \frac{\mu_j}{|\mu_1|+|\mu_2|}(|\mu_1|+|\mu_2|) = 0$ for $j = 1, 2$ as we wanted to prove. \square

4 A representer theorem for the infinite infimal convolution functional

In this section we analyze sparse solutions for inverse problems regularized with the infinite infimal convolution functional introduced in Section 2.2, namely

$$\inf_{\substack{v \in X, \\ \sigma \in \mathcal{M}_+(S)}} \frac{1}{2} \|Av - f\|_Y^2 + \alpha R(v, \sigma), \quad (4.1)$$

where we used the same notations as in Section 2.2.

Sparse solutions of infinite dimensional inverse problems regularized with convex functionals have been recently studied in [7, 10, 49]. In particular, it has been shown that in a general inverse problem, the extremal points of the ball of the convex regularizer can be seen as the atoms of the problem, in analogy with more standard sparsity promoting regularizers, such as the ℓ^1 norm in a finite dimensional setting and the total variation norm for Radon measures. This sentence can be justified by proving so-called representer theorems that, in case of finite-dimensional measurements, ensure the existence of a solution that is representable as a finite linear combination of the mentioned extremal points. In the next theorem, we take advantage of the characterization of extremal points for the ball of the lifted regularizer, see Theorem 3.8, and we use the results in [10] to prove a representer theorem for (4.1) assuming finite-dimensional measurements. In particular, we apply the results in [10] to the lifted problem (3.14) and then, thanks to the equivalence of (3.14) with (4.1), we deduce the representer theorem for (4.1). Note that, in order to apply the results in [10], we need to additionally assume that $J(\cdot, s)$ are seminorms for every $s \in S$. However, we believe that it would be possible to remove this assumption by applying the results in [7].

Theorem 4.1. *Assume that $J(\lambda v, s) = |\lambda|J(v, s)$ for every $\lambda \in \mathbb{R}$, $v \in X$ and $s \in S$. Suppose that the measurement space Y is finite-dimensional with $\dim(Y) = Q \in \mathbb{N}$ and that $\text{con}(\{Av : v \in \bigcup_{s \in S} \text{dom} J(\cdot, s)\}) = Y$, where $\text{con}(C)$ denotes the conical hull of the set C . Then, there exists $(v^*, \sigma^*) \in X \times \mathcal{M}_+(S)$ that minimizes (2.7) such that*

$$v^* = \sum_{i=1}^M \lambda_i v_i, \quad \sigma^* = \sum_{s \in \{s_i: i=1, \dots, M\}} \left\| \sum_{j \in \{1, \dots, M: s_j=s\}} \lambda_j v_j \right\|_X \delta_s, \quad (4.2)$$

where $M \leq Q$, $s_i \in S$, $v_i \in \text{Ext}(B_{s_i}) \setminus \{0\}$ and $\lambda_i > 0$ for every $i = 1, \dots, M$.

Proof. We rewrite (4.1) in $\mathcal{M}(S; X)$ according to (3.14)

$$\inf_{\mu \in \mathcal{M}(S; X)} \frac{1}{2} \|K\mu - f\|_Y^2 + \alpha \int_S J\left(\frac{\mu}{|\mu|}(s), s\right) d|\mu|(s), \quad (4.3)$$

where $K := A \circ T$ and $T\mu := \mu(S)$. Note that $K : \mathcal{M}(S; X) \rightarrow Y$ is a linear operator and it is weak*-to-strong continuous. Indeed, K is weak*-to-weak continuous by Remark A.7, and weak topology and strong topology on Y coincide by the finite dimensionality of Y . Note that $\hat{F}(\lambda\mu) = |\lambda|\hat{F}(\mu)$ for every $\mu \in \mathcal{M}(S; X)$ and $\lambda \in \mathbb{R}$. Indeed, for every $\lambda \neq 0$, Proposition A.2 yields that $\frac{\lambda\mu}{|\lambda\mu|}(s) = \text{sign}(\lambda)\frac{\mu}{|\mu|}(s)$ for $|\mu|$ -a.e. $s \in S$ and hence, using that $J(\lambda v, s) = |\lambda|J(v, s)$ for every $\lambda \in \mathbb{R}$, $v \in X$ and $s \in S$, we have

$$\hat{F}(\lambda\mu) = \int_S J\left(\text{sign}(\lambda)\frac{\mu}{|\mu|}(s), s\right) d(|\lambda\mu|)(s) = |\lambda|\hat{F}(\mu).$$

Therefore, Proposition 3.1 and Theorem 3.2 imply that \hat{F} is a lower semicontinuous seminorm on $\mathcal{M}(S; X)$ equipped with the weak* topology. Moreover, Proposition 3.1 and the Banach–Alaoglu theorem (see also Theorem A.6) imply that the sublevel set of \hat{F} are compact in the weak* topology of $\mathcal{M}(S; X)$.

We now show that $K(\text{dom}\hat{F}) = Y$. Note that it is enough to prove that $Y \subset K(\text{dom}\hat{F}) = A(T(\text{dom}\hat{F}))$. Given $y \in Y$, using that $\text{coni}(\{Av : v \in \bigcup_{s \in S} \text{dom}J(\cdot, s)\}) = Y$, we can write $y = Av$ for $v = \sum_{j=1}^N c_j v_{s_j}$ with $N \in \mathbb{N}$, $s_j \in S$, $v_{s_j} \in \text{dom}J(\cdot, s_j)$ and $c_j \geq 0$. Define then $\mu = \sum_{j=1}^N c_j v_{s_j} \delta_{s_j} \in \mathcal{M}(S; X)$ and note that $\mu \in \text{dom}(\hat{F})$, since by (3.3) in Proposition 3.1, the convexity and the positive one-homogeneity of $J(\cdot, s)$, we have

$$\hat{F}(\mu) = \hat{F}\left(\sum_{j=1}^N c_j v_{s_j} \delta_{s_j}\right) \leq \sum_{j=1}^N c_j J(v_{s_j}, s_j) < \infty. \quad (4.4)$$

We then conclude that $y \in A(T(\text{dom}\hat{F}))$ by observing that $T\mu = v$.

Therefore, we can apply the representer theorem [10, Theorem 3.3] to ensure the existence of a minimizer $\mu^* \in \mathcal{M}(S; X)$ of (4.3) that can be written as

$$\mu^* = \sum_{i=1}^{M'} \lambda_i \mu_i, \quad (4.5)$$

where $M' \leq Q$, $\lambda_i > 0$ and $\mu_i \in \text{Ext}(\{\mu \in \mathcal{M}(S; X) : \hat{F}(\mu) \leq 1\})$ for $i = 1, \dots, M'$. Using Theorem 3.8, we infer that there exist $s_i \in S$ and $v_i \in \text{Ext}(B_{s_i})$ such that $\mu_i = v_i \delta_{s_i}$, so that, by removing from the linear combination (4.5) the measures μ_i that are equal to zero and possibly rearranging the indices, we obtain that

$$\mu^* = \sum_{i=1}^M \lambda_i v_i \delta_{s_i}, \quad (4.6)$$

where $M \leq M'$, $s_i \in S$ and $v_i \in \text{Ext}(B_{s_i}) \setminus \{0\}$. Finally, applying Proposition 3.4 we deduce that the pair $(v^*, \sigma^*) \in X \times \mathcal{M}_+(S)$ defined as

$$v^* = \sum_{i=1}^M \lambda_i v_i, \quad \sigma^* = \sum_{s \in \{s_i : i=1, \dots, M\}} \left\| \sum_{j \in \{1, \dots, M : s_j = s\}} \lambda_j v_j \right\|_X \delta_s$$

is a minimizer of (4.1) as required. \square

5 A generalized conditional gradient method

In this section we will describe how to solve the lifted inverse problem defined in (3.14) by a generalized conditional gradient method. As a consequence of the equivalence provided by Proposition 3.4,

we will then be able to derive a conditional gradient method for (2.7) and (2.8). We recall that the lifted inverse problem consists in minimizing the following functional:

$$\mathcal{G}_\alpha(\mu) = \frac{1}{2} \|A\mu(S) - f\|_Y^2 + \alpha \hat{F}(\mu), \quad (5.1)$$

where

$$\hat{F}(\mu) = \int_S J\left(\frac{\mu}{|\mu|}(s), s\right) d|\mu|(s).$$

As already done in Section 4, denoting by $T : \mathcal{M}(S; X) \rightarrow X$ the linear operator defined as $T\mu := \mu(S)$, we rewrite (5.1) equivalently as

$$\mathcal{G}_\alpha(\mu) = \frac{1}{2} \|K\mu - f\|_Y^2 + \alpha \hat{F}(\mu), \quad (5.2)$$

where $K := A \circ T$. Note again that K is weak*-to-weak continuous by Remark A.7. The pre-adjoint of the linear operator $T : \mathcal{M}(S; X) \rightarrow X$ can be characterized as follows.

Lemma 5.1. *The pre-adjoint of the linear operator T is $T_* : X_* \rightarrow C(S; X_*)$ is given by*

$$(T_*v_*)(s) = v_* \quad \forall s \in S, \quad \forall v_* \in X_*. \quad (5.3)$$

Proof. For every $\mu \in \mathcal{M}(S; X)$ and $v_* \in X_*$ there holds

$$\langle (s \mapsto v_*), \mu \rangle_{\mathcal{M}} = \langle v_*, \mu(S) \rangle_{(X_*, X)} = \langle v_*, T\mu \rangle_{(X_*, X)}, \quad (5.4)$$

showing (5.3). Notice that the first equality in (5.4) follows from the definition of the integral with respect to the vector measure μ . \square

5.1 Description of the algorithm

A generalized conditional gradient algorithm for minimizing \mathcal{G}_α is based on the construction of a sparse iterate $\mu^k \in \mathcal{M}(S; X)$ given by the linear combination of extremal points of B , i.e.,

$$\mu^k = \sum_{i=1}^{N_k} c_i^k \mu_i^k, \quad (5.5)$$

where $N_k \in \mathbb{N}$, $c_i^k > 0$ and $\mu_i^k \in \text{Ext}(B)$ pairwise distinct, for every i . At each iterate $k \in \mathbb{N}$, the algorithm performs two steps: the insertion step and the weights optimization step. In the insertion step a suitable extremal point of B , denoted by $\mu_{N_k+1}^k$, is chosen in such a way that it minimizes a variational problem obtained from (5.1) by linearizing the fidelity term $\frac{1}{2} \|K\mu - f\|_Y^2$ in the previous iterate μ^k . In the weights optimization step the coefficients of the conical combination $\left[\sum_{i=1}^{N_k} c_i^k \mu_i^k \right] + c \mu_{N_k+1}^k$ are optimized with respect to \mathcal{G}_α . The next iterate is then constructed by

$$\mu^{k+1} = \left[\sum_{i=1}^{N_k} \hat{c}_i^k \mu_i^k \right] + \hat{c}_{N_k+1}^k \mu_{N_k+1}^k,$$

where $(\hat{c}_1^k, \dots, \hat{c}_{N_k}^k, \hat{c}_{N_k+1}^k) \in (0, +\infty)^{N_k+1}$ is the solution of the weights optimization step. In the following part of this section we will provide more details on the insertion step and the coefficient optimization step for our problem. Then, we will formulate the generalized conditional gradient algorithm for the lifted inverse problem, see Algorithm 1, and we discuss how to equivalently reformulate this algorithm to compute minimizers of (2.7) and (2.8), see Algorithm 2.

5.1.1 Insertion step

Given a sparse iterate $\mu^k \in \mathcal{M}(S; X)$ that is a conical combination of extremal points of B as in (5.5), we define the dual variable associated to the k -th iteration $p^k \in C(S; X_*)$ as

$$p^k = -K_*(K\mu^k - f). \quad (5.6)$$

Such dual variable p^k represents the negative derivative of the fidelity term in of \mathcal{G}_α defined in (5.2) computed at the point μ^k . We then find the new extremal point of B by solving the linearized problem defined as follows.

Definition 5.2. With p^k as in (5.6) we define the linearized problem in $\mu^k \in \mathcal{M}(S; X)$ as

$$\max_{\mu \in \text{Ext}(B)} \langle p^k, \mu \rangle. \quad (5.7)$$

Note that the linearized problem in Definition 5.2 is, in our setting, equivalent to the linearization of the objective function in a standard Frank–Wolfe algorithm [28]:

$$\max_{\mu \in B} \langle p^k, \mu \rangle. \quad (5.8)$$

Indeed, a simple convexity argument shows that it is always possible to find a maximizer of (5.8) in the set of extremal points of B ; see for example [13, Lemma 3.1]. As a consequence, thanks to Proposition 3.1 and Theorem 3.2, the problem (5.7) admits a solution.

Rephrasing the insertion step as in Definition 5.2 allows to constrain the iterate of our conditional gradient method to be a linear combination of the extremal points of B , highlighting the sparsity promoting structure of the regularizer. As a consequence, we can choose $\mu_{N_k+1}^k \in \arg \max_{\mu \in \text{Ext}(B)} \langle p^k, \mu \rangle$ to be the new extremal point added to the linear combination (5.5).

5.1.2 Weights optimization step

In the insertion step we have selected the new extremal point $\mu_{N_k+1}^k \in \arg \max_{\mu \in \text{Ext}(B)} \langle p^k, \mu \rangle$ to add to the iterate μ^k . The goal of the optimization step is to optimize the coefficients of the sparse conical combination $\left[\sum_{i=1}^{N_k} c_i^k \mu_i^k \right] + c \mu_{N_k+1}^k$ with respect to the energy \mathcal{G}_α :

$$\arg \min_{c_i \in \mathbb{R}_+^{N_k+1}} \frac{1}{2} \left\| \sum_{i=1}^{N_k+1} c_i K \mu_i^k - f \right\|_Y^2 + \alpha \hat{F} \left(\sum_{i=1}^{N_k+1} c_i \mu_i^k \right). \quad (5.9)$$

In practice (see [13, Section 3] for more details), it is enough to optimize the following upper bound of (5.9):

$$\hat{c}^k \in \arg \min_{c_i \in \mathbb{R}_+^{N_k+1}} \frac{1}{2} \left\| \sum_{i=1}^{N_k+1} c_i K \mu_i^k - f \right\|_Y^2 + \alpha \sum_{i=1}^{N_k+1} c_i \hat{F}(\mu_i^k), \quad (5.10)$$

where the bound $\hat{F} \left(\sum_{i=1}^{N_k+1} c_i \mu_i^k \right) \leq \sum_{i=1}^{N_k+1} c_i \hat{F}(\mu_i^k)$ is ensured by the convexity and the positive one-homogeneity of \hat{F} . Note that, since \hat{F} is positively one-homogeneous, we have $\hat{F}(\mu_i^k) \in \{0, 1\}$. Therefore, (3.2) in Theorem 3.1 implies that, if all inserted extremal points are different from zero, then the insertion step can be rewritten as

$$\hat{c}^k \in \arg \min_{c_i \in \mathbb{R}_+^{N_k+1}} \frac{1}{2} \left\| \sum_{i=1}^{N_k+1} c_i K \mu_i^k - f \right\|_Y^2 + \alpha \sum_{i=1}^{N_k+1} c_i. \quad (5.11)$$

Since we cannot exclude that all inserted extremal points are different from zero at all iterations, we will employ the optimization problem in (5.9). However, we refer to Remark 5.4, where we discuss in details how and when the insertion of null extremal points can be avoided.

Remark 5.3. We point out that (5.10) can be a strict upper bound for (5.9) depending on the choice of the functionals $J(\cdot, s)$. To demonstrate this, take $\mu^k = \sum_{i=1}^{N_k+1} c_i^k \mu_i^k$, where $N_k \in \mathbb{N}$, $c_i^k > 0$ and $\mu_i^k \in \text{Ext}(B)$ pairwise distinct, for every i . Thanks to Theorem 3.8, it holds that $\mu_i^k = v_i^k \delta_{s_i^k}$ where $s_i^k \in S$ and $v_i^k \in \text{Ext}(B_{s_i^k})$ for $i = 1, \dots, N_k + 1$. Using Proposition 3.1, we can write

$$\begin{aligned} \hat{F} \left(\sum_{i=1}^{N_k+1} c_i \mu_i^k \right) &= \hat{F} \left(\sum_{s \in \{s_i^k : i=1, \dots, N_k+1\}} \left(\sum_{j \in \{1, \dots, N_k+1 : s_j^k = s\}} c_j v_j^k \right) \delta_s \right) \\ &= \sum_{s \in \{s_i^k : i=1, \dots, N_k+1\}} J \left(\sum_{j \in \{1, \dots, N_k+1 : s_j^k = s\}} c_j v_j^k, s \right). \end{aligned} \quad (5.12)$$

In particular, (5.12) shows that if for every $s \in S$ the functional $J(\cdot, s)$ is linear on conical combinations of elements in $\text{Ext}(B_s)$, then (5.10) is equivalent to (5.9). However, this is in general not true even for simple examples such as $J(v, s) = \|v\|_{L^2}$.

Thanks to the linearity of K , the optimization in (5.10) is a convex quadratic problem in $\mathbb{R}_+^{N_k+1}$ and therefore can be solved efficiently with standard solvers. We then construct the next iterate $\mu^{k+1} \in \mathcal{M}(S; X)$ as $\mu^{k+1} = \sum_{i=1}^{N_k+1} \hat{c}_i^k \mu_i^k$, where \hat{c}_i^k are the optimal weights determined as in (5.9). Algorithm 1 summarizes these steps.

5.1.3 Generalized conditional gradient method for the lifted problem

Algorithm 1 is a generalized conditional gradient method that computes a minimizer of the lifted inverse problem (5.1). Note that we keep track of the extremal points in the representation of the iterate $\mu^k \in \mathcal{M}(S; X)$ in the set \mathcal{A}^k . This is useful because during the weight optimization step some of the optimal $\hat{c}_i^k \in \mathbb{R}_+$ could be zero. If this is the case we want to remove the corresponding extremal point from the linear combination. Such operation is performed by removing from \mathcal{A}^{k+1} the extremal points μ_i^k whose corresponding weight \hat{c}_i^k is equal to zero.

It is possible to add a stopping criterion to Algorithm 1 as in [13, Proposition 3.3]. In particular, we set the algorithm to terminate at the k -th iteration for $k \geq 1$ if

$$\max_{\mu \in \text{Ext}(B)} \langle p^k, \mu \rangle \leq 1 \quad (5.13)$$

or if the newly inserted atom is already among the atoms composing μ^k , i.e., $\mu_{N_k+1}^k \in \mathcal{A}^k$.

Remark 5.4. Note that the stopping criterion (5.13) is met, in particular, if

$$0 \in \arg \max_{\mu \in \text{Ext}(B)} \langle p^k, \mu \rangle. \quad (5.14)$$

Therefore, for all the iterations $k \geq 1$ the algorithm is never inserting 0, since in this case the algorithm would stop. However, it can happen that $\mu = 0$ is inserted in the first iteration, since the stopping criterion is not given for $k = 0$. It is possible to circumvent this issue in several ways. First, Theorem 3.8 states that $0 \notin \text{Ext}(B)$ in case there exists $s \in S$ such that $0 \notin \text{Ext}(B_s)$; if this happens, no workarounds are needed. Second, if we initialize Algorithm 1 with the zero measure, then $p^0 = K_* f$ and thus, $0 \in \arg \max_{\mu \in \text{Ext}(B)} \langle p^0, \mu \rangle$ implies that 0 is a minimizer of (5.2) due to general optimality conditions derived, for example, in [13, Proposition 2.3] and the equivalence between (5.7) and (5.8) [13, Lemma 3.1]. Third, it is possible to perform an additional weight optimization step for the initial measure $\mu^0 = \sum_{i=1}^{N_0} c_i^0 \mu_i^0$, before the first insertion step; in this case the stopping criterion can be also given for the first iteration $k = 0$ and the insertion of zero would cause the algorithm to stop.

Algorithm 1 Generalized conditional gradient method for the lifted inverse problem

Initialize: $\mu^0 = \sum_{i=1}^{N_0} c_i^0 \mu_i^0$ for $\mu_i^0 \in \text{Ext}(B) \setminus \{0\}$ pairwise distinct, $c_i^0 > 0$ for $i = 1, \dots, N_0$,
 $\mathcal{A}^0 = \{\mu_1^0, \dots, \mu_{N_0}^0\}$
for $k = 0, 1, 2 \dots$ **do**
 $p^k \leftarrow -K_*(K\mu^k - f)$
 $\mu_{N_k+1}^k \in \arg \max_{\mu \in \text{Ext}(B)} \langle p^k, \mu \rangle$
 if $k \geq 1$, $\langle p^k, \mu_{N_k+1}^k \rangle \leq 1$ or $\mu_{N_k+1}^k \in \mathcal{A}^k$ **then**
 return μ^k
 end if
 $\hat{c}^k \in \arg \min_{c \in \mathbb{R}_+^{N_k+1}} \frac{1}{2} \|\sum_{i=1}^{N_k+1} c_i K \mu_i^k - f\|_Y^2 + \alpha \sum_{i=1}^{N_k+1} c_i \hat{F}(\mu_i^k)$
 $\mu^{k+1} \leftarrow \sum_{i=1}^{N_k+1} \hat{c}_i^k \mu_i^k$
 $\mathcal{A}^{k+1} \leftarrow \{\mu_i^k : \hat{c}_i^k > 0\}$, $N_{k+1} \leftarrow \#\mathcal{A}^{k+1}$
 $k \leftarrow k + 1$
end for
return μ^k

5.1.4 Generalized conditional gradient method for the infinite infimal convolution functional

Using the characterization of extremal points of B from Theorem 3.8 together with Lemma 5.1, we can rewrite Algorithm 1 in the space $X \times \mathcal{M}_+(S)$ to obtain an implementable reformulation and an iterate that converges to a minimizer of (2.7), see Section 5.2. By Theorem 3.8 the iterate μ^k can be written as

$$\mu^k = \sum_{i=1}^{N_k} c_i^k v_i^k \delta_{s_i^k}, \quad (5.15)$$

where $N_k \in \mathbb{N}$, $c_i^k > 0$, $s_i^k \in S$ and $v_i^k \in \text{Ext}(B_{s_i^k})$ for $i = 1, \dots, N_k$.

Remark 5.5. If there exists $s \in S$ such that $0 \notin \text{Ext}(B_s)$, then Theorem 3.8 implies that $0 \notin \text{Ext}(B)$ and thus v_i^k belongs to $\text{Ext}(B_{s_i^k}) \setminus \{0\}$ for $i = 1, \dots, N_k$.

By applying Lemma 5.1, the dual variable $p^k \in C(S; X_*)$ defined in (5.6) can be rewritten as

$$p^k(s) = -A_*(AT\mu^k - f) \quad \forall s \in S, \quad (5.16)$$

and thus, applying the definition of T we get

$$p^k(s) = -A_* \left(\sum_{i=1}^{N_k} c_i^k A v_i^k - f \right) \quad \forall s \in S. \quad (5.17)$$

We choose a new extremal point of B , denoted by $v_{N_k+1}^k \delta_{s_{N_k+1}^k}$, to be a conical combination (5.15) that minimizes the linearized problem defined in Definition 5.2. More precisely, using (5.17), the insertion step can be written as follows

$$\max_{\mu \in \text{Ext}(B)} \langle p^k, \mu \rangle = \max_{\substack{(v,s) \in X \times S, \\ v \in \text{Ext}(B_s)}} \langle p^k, v \delta_s \rangle = \max_{\substack{(v,s) \in X \times S, \\ v \in \text{Ext}(B_s)}} \langle -A_* \left(\sum_{i=1}^{N_k} c_i^k A v_i^k - f \right), v \rangle.$$

Therefore, in the insertion step we seek to solve

$$(v_{N_k+1}^k, s_{N_k+1}^k) \in \arg \max_{\substack{(v,s) \in X \times S, \\ v \in \text{Ext}(B_s)}} \langle -A_* \left(\sum_{i=1}^{N_k} c_i^k A v_i^k - f \right), v \rangle. \quad (5.18)$$

The weights optimization step can be also rewritten as the quadratic problem

$$\hat{c}^k \in \arg \min_{c \in \mathbb{R}_+^{N_k+1}} \frac{1}{2} \left\| \sum_{i=1}^{N_k+1} c_i A v_i^k - f \right\|_Y^2 + \alpha \sum_{i=1}^{N_k+1} c_i J(v_i^k, s_i^k) \quad (5.19)$$

and the next iterate $\mu^{k+1} \in \mathcal{M}(S; X)$ is produced as $\mu^{k+1} = \sum_{i=1}^{N_k+1} \hat{c}_i^k v_i^k \delta_{s_i^k}$.

In order to rewrite Algorithm 1 in $X \times \mathcal{M}_+(S)$, we define the iterate $(v^k, \sigma^k) \in X \times \mathcal{M}_+(S)$ as $v^k = \sum_{i=1}^{N_k} c_i^k v_i^k$ and $\sigma^k = \sum_{s \in \{s_i^k: i=1, \dots, N_k\}} \left\| \sum_{j \in \{1, \dots, N_k: s_j^k=s\}} c_j^k v_j^k \right\|_X \delta_s$. In Algorithm 2 we summarize the generalized conditional gradient algorithm we use to solve (5.2). Similarly to Algorithm 1, we keep track of the pairs (v_i^k, s_i^k) that represent v^k and σ^k by defining the sets $\Xi^k = \{(v_1^k, s_1^k), \dots, (v_{N_k}^k, s_{N_k}^k)\}$. At each iteration we add the solution of (5.18) to Ξ^k and, after the weights optimization step, we remove the pairs (v_i^k, s_i^k) for which the corresponding weights \hat{c}_i^k are equal to zero.

Thanks to Proposition 3.4 and Theorem 3.8, the stopping criterion we used in Algorithm 1 can be reformulated for Algorithm 2 as follows. We set the algorithm to terminate at the k -th iteration, for $k \geq 1$, if

$$\arg \max_{\substack{(v,s) \in X \times S, \\ v \in \text{Ext}(B_s)}} \langle p^k, v \rangle \leq 1. \quad (5.20)$$

Remark 5.6. If there exists $s \in S$ such that $0 \notin \text{Ext}(B_s)$, then Theorem 3.8 implies that $0 \notin \text{Ext}(B)$. Thus, the insertion step in (5.18) can be rewritten as

$$(v_{N_k+1}^k, s_{N_k+1}^k) \in \arg \max_{\substack{(v,s) \in X \times S, \\ v \in \text{Ext}(B_s) \setminus \{0\}}} \left\langle -A_* \left(\sum_{i=1}^{N_k} c_i^k A v_i^k - f \right), v \right\rangle \quad (5.21)$$

and the stopping criterion (5.20) can be rewritten as

$$\arg \max_{\substack{(v,s) \in X \times S, \\ v \in \text{Ext}(B_s) \setminus \{0\}}} \langle p^k, v \rangle \leq 1. \quad (5.22)$$

If $0 \in \text{Ext}(B_s)$ for every $s \in S$, then the pair $(0, s)$ can be inserted by Algorithm 2 in the first iteration, since the stopping criterion is not given for $k = 0$. Following the same arguments of Remark 5.4, it is possible to avoid the insertion of the pair $(0, s)$ if we initialize Algorithm 2 with $N_0 = 0$ and $\Xi = \{\}$. Indeed, in this case, the insertion of $(0, s)$ for $s \in S$ implies that the zero measure is a solution of (3.14) and thus the pair $(0, 0)$ is a solution of (2.7) by Proposition 3.4. In case of general initializations $v^0 = \sum_{i=1}^{N_0} c_i^0 v_i^0$ and $\sigma^0 = \sum_{s \in \{s_i^0: i=1, \dots, N_0\}} \left\| \sum_{j \in \{1, \dots, N_0: s_j^0=s\}} c_j^0 v_j^0 \right\|_X \delta_s$ one can resort to a preliminary weight optimization step to extend the stopping criterion to the first iteration as in Remark 5.4.

In the remaining part of the paper and in the description of the algorithms we will not discuss anymore the implementation choices necessary to avoid the insertion of the extremal point 0, but we stress that in some instances it is important to treat this case with care; see for example [12].

5.2 Sublinear convergence

The goal of this section is prove sublinear convergence for the iterates generated by Algorithms 1 and 2. To this end, we define the functional distance associated to the objective functional \mathcal{G}_α , see (5.1), as

$$r_{\mathcal{G}_\alpha}(\mu) := \mathcal{G}_\alpha(\mu) - \min_{\tilde{\mu} \in \mathcal{M}(S; X)} \mathcal{G}_\alpha(\tilde{\mu}) \quad (5.23)$$

Algorithm 2 Generalized conditional gradient method for infinite infimal convolutions

Initialize: $v^0 = \sum_{i=1}^{N_0} c_i^0 v_i^0$, $\sigma^0 = \sum_{s \in \{s_i^0: i=1, \dots, N_0\}} \left\| \sum_{j \in \{1, \dots, N_0: s_j^0=s\}} c_j^0 v_j^0 \right\|_X \delta_s$
 $\Xi^0 = \{(v_1^0, s_1^0), \dots, (v_{N_0}^0, s_{N_0}^0)\}$ with (v_i^0, s_i^0) pairwise distinct, $c_i^0 > 0$, $s_i^0 \in S$, $v_i^0 \in \text{Ext}(B_{s_i^0}) \setminus \{0\}$
 for $i = 1, \dots, N_0$.
for $k = 0, 1, 2 \dots$ **do**
 $p^k \leftarrow -A_* (Av^k - f)$
 $(v_{N_{k+1}}^k, s_{N_{k+1}}^k) \in \arg \max_{\substack{(v,s) \in X \times S, \\ v \in \text{Ext}(B_s)}} \langle p^k, v \rangle$
 if $k \geq 1$, $\langle p^k, v_{N_{k+1}}^k \rangle \leq 1$ or $(v_{N_{k+1}}^k, s_{N_{k+1}}^k) \in \Xi^k$ **then**
 return (v^k, σ^k)
 end if
 $(\hat{c}_1^k, \dots, \hat{c}_{N_{k+1}}^k) \in \arg \min_{c \in \mathbb{R}_+^{N_{k+1}}} \frac{1}{2} \left\| \sum_{i=1}^{N_{k+1}} c_i A v_i^k - f \right\|_Y^2 + \alpha \sum_{i=1}^{N_{k+1}} c_i J(v_i^k, s_i^k)$
 $v^{k+1} \leftarrow \sum_{i=1}^{N_{k+1}} \hat{c}_i^k v_i^k$, $\sigma^{k+1} \leftarrow \sum_{s \in \{s_i^k: i=1, \dots, N_{k+1}\}} \left\| \sum_{j \in \{1, \dots, N_{k+1}: s_j^k=s\}} \hat{c}_j^k v_j^k \right\|_X \delta_s$
 $\Xi^{k+1} \leftarrow \{(v_i^k, s_i^k) \in \Xi^k \cup \{(v_{N_{k+1}}^k, s_{N_{k+1}}^k)\} : \hat{c}_i^k > 0\}$
 $N_{k+1} \leftarrow \#\Xi^{k+1}$, $k \leftarrow k + 1$
end for
return (v^k, σ^k)

for all $\mu \in \mathcal{M}(S; X)$ and the functional distance associated to the original inverse problem (2.7) as

$$r_{\mathcal{F}_\alpha}(v, \sigma) := \mathcal{F}_\alpha(v, \sigma) - \min_{\substack{\tilde{v} \in X, \\ \tilde{\sigma} \in \mathcal{M}_+(S)}} \mathcal{F}_\alpha(\tilde{v}, \tilde{\sigma}) \quad (5.24)$$

for all $v \in X$ and $\sigma \in \mathcal{M}_+(S)$, where

$$\mathcal{F}_\alpha(v, \sigma) := \frac{1}{2} \|Av - f\|_Y^2 + \alpha R(v, \sigma).$$

The following convergence result for $r_{\mathcal{G}_\alpha}$ holds.

Theorem 5.7. *Suppose that hypotheses (H1), (H2) and (H3) hold. Additionally assume that*

- (I1) *for every sequence $\{v^k\}_k$ in X such that there exists suitable $\sigma^k \in \mathcal{M}_+(S)$ for which $v^k \in \text{dom}(R(\cdot, \sigma^k))$ and $v^k \xrightarrow{*} v$, $v \in X$, it holds that $\lim_{k \rightarrow +\infty} Av^k = Av$.*

Let $\{\mu^k\}_k$ be a sequence generated by Algorithm 1. Then either Algorithm 1 terminates after a finite number of steps returning a minimizer of \mathcal{G}_α , or there exists $C > 0$ depending only on f such that

$$r_{\mathcal{G}_\alpha}(\mu^k) \leq \frac{C}{k+1}, \quad \text{for all } k \in \mathbb{N}. \quad (5.25)$$

Moreover, each accumulation point μ^ of $\{\mu^k\}_k$ with respect to the weak* topology of $\mathcal{M}(S; X)$ is a minimizer for \mathcal{G}_α . If \mathcal{G}_α admits a unique minimizer μ^* , then $\mu^k \xrightarrow{*} \mu^*$ along the whole sequence.*

Proof. The proof of Theorem 5.7 proceeds by specializing [13, Theorem 3.4] in our setting for the generalized conditional gradient method applied to the functional \mathcal{G}_α . We need to verify that the assumptions (A1), (A2), (A3) provided in Section 2 in [13] hold in our setting. First note that $\mathcal{M}(S; X)$ is a Banach space when equipped with the total variation norm and $C(S; X_*)$ is its predual (see Theorem A.6). Additionally, the functional $\hat{F} : \mathcal{M}(S; X) \rightarrow [0, \infty]$ is convex, positively one-homogeneous, weakly-* lower semicontinuous and coercive in the sense that its sublevel sets are weakly-* compact (see Proposition 3.1 and Theorem 3.2). This ensures the validity of (A1) and (A2). It remains to prove that the linear operator $K = A \circ T$, where $T\mu = \mu(S)$, is weak*-to-weak continuous and it satisfies assumption (A3) in [13] as a consequence of (I1). Applying Remark A.7, we immediately deduce that K is weak*-to-weak continuous since A is weak*-to-weak continuous. Moreover, given a sequence $\{\mu^n\}_n$ in $\mathcal{M}(S; X)$ such that $\mu^n \in \text{dom}(\hat{F})$ we

have that $\mu^n(S) \in \text{dom}(R(\cdot, |\mu^n|))$. Indeed, by the definition of $R(\mu^n(S), |\mu^n|)$ and the fact that $\mathcal{I}_{|\mu^n|} \frac{\mu^n}{|\mu^n|} = \mu^n(S)$ we have that $R(\mu^n(S), |\mu^n|) \leq \hat{F}(\mu^n) < \infty$. Therefore, considering a sequence $\{\mu^n\}_n$ in $\text{dom}(\hat{F})$ such that $\mu^n \xrightarrow{*} \mu$, using Assumption (I1) and recalling Remark A.7, we conclude that $K\mu^n \rightarrow K\mu$ strongly. We can therefore apply Theorem 3.4 in [13] and conclude the proof. \square

By applying Proposition 3.4 and Theorem 5.7, we can deduce from the previous theorem a convergence result for Algorithm 2 as follows.

Corollary 5.8. *Let*

$$v^k = \sum_{i=1}^{N_k} c_i^k v_i^k \in X, \quad \sigma^k = \sum_{s \in \{s_i^k: i=1, \dots, N_k\}} \left\| \sum_{j \in \{1, \dots, N_k: s_j^k=s\}} c_j^k v_j^k \right\|_X \delta_s \quad (5.26)$$

be the iterates generated by Algorithm 2. If Algorithm 2 terminates after a finite number of iterations returning $(v^*, \sigma^*) \in X \times \mathcal{M}_+(S)$, then (v^*, σ^*) is a minimizer of \mathcal{F}_α . Moreover, there exists $C > 0$ depending only on f such that

$$r_{\mathcal{F}_\alpha}(v^k, \sigma^k) \leq \frac{C}{k+1}, \quad \text{for all } k \in \mathbb{N}. \quad (5.27)$$

Proof. If Algorithm 2 terminates after a finite number of steps returning (v^*, σ^*) , the proof follows from Theorem 5.7 by noticing that, given a minimizer for \mathcal{G}_α denoted by $\mu^* \in \mathcal{M}(S; X)$, Proposition 3.4 ensures that $(\frac{\mu^*}{|\mu^*|}, |\mu^*|)$ is a minimizer for (2.8) and thus, $(\mu^*(S), |\mu^*|)$ is a minimizer for (2.7). Now note that

$$\mathcal{F}_\alpha(v, \sigma) \leq \mathcal{G}_\alpha(\mu)$$

for every μ such that $\mu(S) = v$ and $|\mu| = \sigma$. Therefore, applying Proposition 3.4, we deduce that

$$r_{\mathcal{F}_\alpha}(v^k, \sigma^k) \leq r_{\mathcal{G}_\alpha}(\mu^k). \quad (5.28)$$

An application of Theorem 5.7 concludes the proof. \square

6 Illustrative applications and numerics

The goal of this section is to provide illustrative examples of inverse problems with infinite infimal convolution regularization and to show that the generalized conditional gradient method introduced in Section 5 can be employed to solve them efficiently without discretizing the set S . In Section 6.1 we consider the infimal convolution of fractional-Laplacian-type functionals with different fractional orders. In Section 6.2 we focus on the infimal convolution of anisotropic fractional-Laplacian-type regularizers.

All of the experiments in this paper are carried out on Python3 on a MacBook Pro with 8 GB RAM and an Intel®Core™ i5, Quad-Core, 2.3 GHz.

6.1 Learning the orders of 1-dimensional fractional Laplacians in an infinite infimal convolution

In this section we apply our setting to an inverse problem regularized by the infimal convolution of fractional-Laplacian-type operators. We refer to Appendix B for basic definitions and results regarding L^2 -periodic functions and fractional-Laplacian-type operators needed in this section. Our goal is to define an infinite infimal convolution of functionals that regularizes a given observation by penalizing combinations of fractional order derivatives, where the orders themselves are learnt and depend on the frequencies of the observation. The higher the frequency of the observation, the lower the order of the regularization applied to that frequency. We use this regularizer to denoise periodic signals. The goal of this example is to assess the performance of our regularization functional in a simple setting and to see how the optimization procedure takes advantage of sparsity in the space of parameters.

6.1.1 Setting

We denote by T^1 the 1-dimensional torus and consider $S = [0, 1]$ and $X = L^2_\circ(T^1)$ endowed with the L^2 -norm. Here, $L^2_\circ(T^1)$ denotes the space of L^2 -functions on T^1 with zero mean, see Appendix B.1. Notice that X is a Hilbert space, being a closed subset of $L^2(T^1)$, and therefore self-dual by the Riesz representation theorem. We define the functional $J : L^2_\circ(T^1) \times [0, 1] \rightarrow [0, +\infty]$ as follows

$$J(v, s) := \begin{cases} \frac{1}{s^\eta} \sqrt{\sum_{m \in \mathbb{Z}} |m|^{4s} |\hat{v}(m)|^2} & \text{if } s > 0, \\ 0 & \text{if } v = 0, s = 0, \\ +\infty & \text{otherwise,} \end{cases} \quad (6.1)$$

where $\eta > 0$ is a fixed parameter. The infinite infimal convolution of the functionals $J(\cdot, s)$ has the desired effect of penalizing lower regularity. Indeed, due to the factor $1/s^\eta$ in the definition of J , a higher smoothness in the observation receives a lower weight in the regularization.

Remark 6.1. We remark that if $s > 0$ and $v \in H^{2s}(T^1)$, then $J(v, s) = \frac{1}{s^\eta} \frac{1}{\sqrt{2\pi}} \|(-\Delta^s)v\|_{L^2(T^1)}$ is the rescaled L^2 -norm of the s -fractional Laplacian (see Remark B.4).

We consider the following regularizer given by an infinite infimal convolution of $J(\cdot, s)$

$$R(v, \sigma) = \inf_{u \in L^1((S, \sigma); L^2_\circ(T^1))} \int_0^1 J(u(s), s) d\sigma(s) + \mathbb{1}_{\{v\}}(\mathcal{I}_\sigma u)$$

and the corresponding denoising problem

$$\inf_{\substack{v \in L^2_\circ(T^1), \\ \sigma \in \mathcal{M}_+(S)}} \frac{1}{2} \|v - f\|_{L^2(T^1)}^2 + \alpha R(v, \sigma), \quad (6.2)$$

where $f \in L^2(T^1)$ is a given observation and $\alpha > 0$. We now verify that this problem satisfies the assumptions made in Section 2.2.

Lemma 6.2. *The functional J defined in (6.1) satisfies (H1), (H2) and (H3) in Section 2.2.*

Proof. Assumption (H1) is straightforward from the definition of $J(v, s)$ in (6.1). Let us show Assumption (H2). Let $\{(v^n, s_n)\}_n$ be a sequence in $L^2_\circ(T^1) \times [0, 1]$ such that $v^n \rightharpoonup v$ in $L^2_\circ(T^1)$ and $s_n \rightarrow s$. Without loss of generality we can assume that

$$\liminf_{n \rightarrow +\infty} J(v^n, s_n) = \lim_{n \rightarrow +\infty} J(v^n, s_n) < \infty. \quad (6.3)$$

In particular, for all but finitely many n either $s_n > 0$ or $(v^n, s_n) = (0, 0)$. Since, if $(v^n, s_n) = (0, 0)$ for all but finitely many the statement is clear, we can suppose without loss of generality that $s_n > 0$ for every n . Note that $\lim_{n \rightarrow +\infty} |m|^{2s_n} \hat{v}^n(m) = |m|^{2s} \hat{v}(m)$ for every $m \in \mathbb{Z}$. Indeed, for every $m \in \mathbb{Z}$ it holds

$$\lim_{n \rightarrow +\infty} |m|^{2s_n} \hat{v}^n(m) = \lim_{n \rightarrow +\infty} \frac{|m|^{2s_n}}{2\pi} \int_0^{2\pi} e^{-imx} v^n(x) dx = \frac{|m|^{2s}}{2\pi} \int_0^{2\pi} e^{-imx} v(x) dx = |m|^{2s} \hat{v}(m).$$

Therefore, if $s > 0$, by Fatou's Lemma we have

$$\liminf_{n \rightarrow +\infty} J(v^n, s_n) = \liminf_{n \rightarrow +\infty} \frac{1}{s_n^\eta} \sqrt{\sum_{m \in \mathbb{Z}} |m|^{4s_n} |\hat{v}^n(m)|^2} \geq \frac{1}{s^\eta} \sqrt{\sum_{m \in \mathbb{Z}} |m|^{4s} |\hat{v}(m)|^2} = J(v, s).$$

In the case that $s = 0$, (6.3) implies

$$0 \leq \lim_{n \rightarrow +\infty} \sqrt{\sum_{m \in \mathbb{Z}} |m|^{4s_n} |\hat{v}^n(m)|^2} \leq \left(\sup_{n \in \mathbb{N}} J(v^n, s_n) \right) \left(\lim_{n \rightarrow +\infty} s_n^\eta \right) = 0,$$

and consequently $\lim_{n \rightarrow +\infty} \sum_{m \in \mathbb{Z}} |\hat{v}^n(m)|^2 = 0$. Since $v^n \rightharpoonup v$ we infer that $v = 0$ and $J(v, s) = 0$, showing that $\liminf_{n \rightarrow +\infty} J(v^n, s_n) \geq J(v, s)$.

Finally, let us verify Assumption (H3). If $s > 0$ and $v \in H^{2s}(T^1)$, then

$$J(v, s) = \frac{1}{s^\eta} \frac{1}{\sqrt{2\pi}} \|(-\Delta^s)v\|_{L^2(T^1)} \geq \frac{1}{\sqrt{2\pi}} \|v\|_{L^2(T^1)}, \quad (6.4)$$

thanks to Remark B.4 and the Poincaré inequality in Theorem B.5. If instead $v \in L^2(T^1) \setminus H^{2s}(T^1)$ or $s = 0$, the inequality $J(v, s) \geq \frac{1}{\sqrt{2\pi}} \|v\|_{L^2(T^1)}$ is immediate from the definition of $J(v, s)$. \square

6.1.2 The generalized conditional gradient method

We now design a generalized conditional gradient method to solve (6.2) following the general procedure outlined in Section 5. Define $B_s = \{v \in L^2_\circ(T^1) : \alpha J(v, s) \leq 1\}$. Lemma B.6 with the choice $f(m) = \alpha^2 |m|^{4s}$ provides the following characterization of the extremal points of B_s :

$$\text{Ext}(B_s) = \left\{ v \in L^2(T^1) : \alpha \frac{1}{s^\eta} \sqrt{\sum_{m \in \mathbb{Z}} |m|^{4s} |\hat{v}(m)|^2} = 1 \right\},$$

where, in particular, we observe that $0 \notin \text{Ext}(B_s)$ for $s \in [0, 1]$, cf. Remark 5.4. Therefore, given an iterate $\mu^k = \sum_{i=1}^{N_k} c_i^k v_i^k \delta_{s_i^k}$ for $N_k \in \mathbb{N}$, $c_i^k > 0$, $v_i^k \in \text{Ext}(B_s)$ and $s_i^k \in [0, 1]$, the insertion step in (5.18) amounts to solving

$$\arg \max_{\substack{s \in S, \\ v \in L^2_\circ(T^1)}} \left\langle f - \sum_{i=1}^{N_k} c_i^k v_i^k, v \right\rangle_{L^2(T^1)} + \mathbb{1}_{\left\{ \frac{\alpha}{s^\eta} \sqrt{\sum_{m \in \mathbb{Z}} |m|^{4s} |\hat{v}(m)|^2} = 1 \right\}}(v). \quad (6.5)$$

The following lemma provides an explicit formula for the inserted $v \in L^2_\circ(T^1)$ for a fixed $s \in [0, 1]$.

Lemma 6.3. *Given $w \in L^2_\circ(T^1)$ and $s \in [0, 1]$, consider the variational problem*

$$\max_{v \in L^2_\circ(T^1)} \langle w, v \rangle_{L^2(T^1)} + \mathbb{1}_{\left\{ \frac{\alpha}{s^\eta} \sqrt{\sum_{m \in \mathbb{Z}} |m|^{4s} |\hat{v}(m)|^2} = 1 \right\}}(v). \quad (6.6)$$

A solution of (6.6) can be computed as

$$v^*(x) = \frac{1}{A} \sum_{m \in \mathbb{Z} \setminus \{0\}} |m|^{-4s} \hat{w}(m) e^{im \cdot x}, \quad (6.7)$$

where $A = \frac{\alpha}{s^\eta} \sqrt{\sum_{m \in \mathbb{Z} \setminus \{0\}} |m|^{-4s} |\hat{w}(m)|^2}$.

Proof. First note that existence of a minimizer for (6.6) follows by specializing the arguments of Section 5.1.1 where a general insertion step is discussed. Since $w \in L^2_\circ(T^1)$, we can represent w and every $v \in L^2_\circ(T^1)$ as

$$w(x) = \sum_{m \in \mathbb{Z} \setminus \{0\}} \hat{w}(m) e^{im \cdot x}, \quad v(x) = \sum_{m \in \mathbb{Z} \setminus \{0\}} \hat{v}(m) e^{im \cdot x}. \quad (6.8)$$

Therefore, with Plancherel's formula (B.7), the variational problem (6.6) can be rewritten equivalently in the Fourier domain as

$$\arg \max_{\hat{v} \in \ell^2(\mathbb{Z})} \langle \hat{w}, \hat{v} \rangle_{\ell^2(\mathbb{Z})} \quad \text{s.t.} \quad \frac{\alpha^2}{s^{2\eta}} \sum_{m \in \mathbb{Z}} |m|^{4s} |\hat{v}(m)|^2 = 1. \quad (6.9)$$

By the substitutions

$$\tilde{w}(m) = \hat{w}(m) \frac{s^\eta}{\alpha} |m|^{-2s} \quad \forall m \in \mathbb{Z} \setminus \{0\} \text{ and } \tilde{w}(0) = 0, \quad (6.10)$$

$$\tilde{v}(m) = \hat{v}(m) \frac{\alpha}{s^\eta} |m|^{2s} \quad \forall m \in \mathbb{Z}, \quad (6.11)$$

we can rewrite (6.9) as $\arg \max_{\tilde{v} \in \ell^2(\mathbb{Z})} \langle \tilde{w}, \tilde{v} \rangle_{\ell^2(\mathbb{Z})}$ under the constraint $\|\tilde{v}(m)\|_{\ell^2(\mathbb{Z})} = 1$ that has the unique solution $\tilde{v}^*(m) = \tilde{w}(m)/\|\tilde{w}\|_{\ell^2(\mathbb{Z})}$. Therefore, by using (6.10), (6.11) and the inverse Fourier transform, a simple computation leads to (6.7). \square

In Algorithm 3 we rewrite the generalized conditional gradient method specialized for (6.2). We initialize the algorithm with a tuple of pairs in $X \times S$ denoted by $\Xi^0 = \{(v_1^0, s_1^0), \dots, (v_{N_0}^0, s_{N_0}^0)\}$ where $N_0 \in \mathbb{N}$, $s_i^0 \in S$ and $v_i^0 \in \text{Ext}(B_{s_i^0})$, and a tuple of positive weights $(c_1^0, \dots, c_{N_0}^0)$. In the insertion step at the k -th iteration of the algorithm, the newly inserted atom is computed by first applying Lemma 6.3 with $w = f - \sum_{i=1}^{N_k} c_i^k v_i^k$ to find the optimal solution for a fixed $s \in [0, 1]$, denoted by $v_{N_{k+1}}^k(s)$. Then, $v_{N_{k+1}}^k(s)$ is optimized with respect to s by solving

$$\max_{s \in [0, 1]} \langle p^k, v_{N_{k+1}}^k(s) \rangle. \quad (6.12)$$

We remark that the computation of (6.12) is performed using a Python implementation of a basin-hopping-type algorithm whose performance is good enough to compute $\arg \max_{s \in [0, 1]} \langle p^k, v_{N_{k+1}}^k(s) \rangle$ efficiently and with satisfactory accuracy.

Algorithm 3 Generalized conditional gradient method for the infinite infimal convolution of fractional Laplacians with adaptive order

Initialize: $v^0 = \sum_{i=1}^{N_0} c_i^0 v_i^0$, $\sigma^0 = \sum_{s \in \{s_i^0: i=1, \dots, N_0\}} \left\| \sum_{j \in \{1, \dots, N_0: s_j^0=s\}} c_j^0 v_j^0 \right\|_X \delta_s$
 $\Xi^0 = \{(v_1^0, s_1^0), \dots, (v_{N_0}^0, s_{N_0}^0)\}$ with (v_i^0, s_i^0) pairwise distinct, $c_i^0 > 0$, $s_i^0 \in S$, $v_i^0 \in \text{Ext}(B_{s_i^0})$ for $i = 1, \dots, N_0$.
for $k = 0, 1, 2 \dots$ **do**
 $p^k \leftarrow f - \sum_{i=1}^{N_k} c_i^k v_i^k$
 $v_{N_{k+1}}^k(s)(x) = \frac{s^\eta}{\alpha \sqrt{\sum_{m \in \mathbb{Z} \setminus \{0\}} |m|^{-4s} |\hat{p}^k(m)|^2}} \sum_{m \in \mathbb{Z} \setminus \{0\}} |m|^{-4s} \hat{p}^k(m) e^{im \cdot x}$
 $s_{N_{k+1}}^k \in \arg \max_{s \in [0, 1]} \langle p^k, v_{N_{k+1}}^k(s) \rangle$
 $v_{N_{k+1}}^k \leftarrow v_{N_{k+1}}^k(s_{N_{k+1}}^k)$
if $k \geq 1$, $\langle p^k, v_{N_{k+1}}^k \rangle \leq 1$ or $(v_{N_{k+1}}^k, s_{N_{k+1}}^k) \in \Xi^k$ **then**
 return (v^k, σ^k)
end if
 $(\hat{c}_1^k, \dots, \hat{c}_{N_{k+1}}^k) \in \arg \min_{c \in \mathbb{R}_+^{N_{k+1}}} \frac{1}{2} \left\| \sum_{i=1}^{N_{k+1}} c_i v_i^k - f \right\|_{L^2(T^1)}^2 + \alpha \sum_{i=1}^{N_{k+1}} c_i$
 $v^{k+1} \leftarrow \sum_{i=1}^{N_{k+1}} \hat{c}_i^k v_i^k$, $\sigma^{k+1} \leftarrow \sum_{s \in \{s_i^k: i=1, \dots, N_{k+1}\}} \left\| \sum_{j \in \{1, \dots, N_{k+1}: s_j^k=s\}} \hat{c}_j^k v_j^k \right\|_X \delta_s$
 $\Xi^{k+1} \leftarrow \{(v_i^k, s_i^k) \in \Xi^k \cup \{(v_{N_{k+1}}^k, s_{N_{k+1}}^k)\} : \hat{c}_i^k > 0\}$
 $N_{k+1} \leftarrow \#\Xi^{k+1}$, $k \leftarrow k + 1$
end for
return (v^k, σ^k)

6.1.3 Numerical experiments

Consider a signal constructed in the following way. Define

$$v(x) = \frac{1}{2\pi} \sum_{m \in \mathbb{Z} \setminus \{0\}} \hat{v}(m) e^{im \cdot x}, \quad (6.13)$$

where $\hat{v}(m)$ is such that $\hat{v}(3) = \hat{v}(-3) = 8$, $\hat{v}(20) = \hat{v}(-20) = 2$ and $\hat{v}(m) = 0$ otherwise. We then construct the data f by corrupting v with Gaussian noise of standard deviation $7 \cdot 10^{-2}$. In the definition of J and in the inverse problem in (6.2) we make the parameter choices, $\eta = 2$, $\alpha = 1.5 \cdot 10^{-3}$, and we initialize Algorithm 3 with $c^0 = 1$ and $\Xi^0 = \{(\frac{v^0}{J(1, v^0)}, 1)\}$, where $v^0(x) = 3\sqrt{\frac{2}{\pi}} \cos(3x)$.

Figure 1 shows the reconstruction obtained by Algorithm 3 and Figure 2 shows the orders $s \in [0, 1]$ selected by the conditional gradient method and a convergence graph where we plot an approximation of the residual $r_{\mathcal{F}_\alpha}$ (5.24) at each iteration in logarithmic scale. More precisely, denoting by $(v^k, \sigma^k) = (\sum_{i=1}^{N_k} c_i^k v_i^k, \sum_{s \in \{s_i^k: i=1, \dots, N_k\}} \|\sum_{j \in \{1, \dots, N_k: s_j^k=s\}} c_j^k v_j^k\|_X \delta_s)$ the iterate generated by Algorithm 3 and by $(v^{\bar{k}}, \sigma^{\bar{k}})$ the output, we plot the approximate residual $\hat{r}_{\mathcal{F}_\alpha}(\mu_k)$ defined as

$$\hat{r}_{\mathcal{F}_\alpha}(\mu_k) = \frac{1}{2} \|v^k - f\|_{L^2(T^1)}^2 + \alpha \sum_{i=1}^{N_k} c_i^k J(v_i^k, s_i^k) - \frac{1}{2} \|v^{\bar{k}} - f\|_{L^2(T^1)}^2 - \alpha \sum_{i=1}^{N_{\bar{k}}} c_i^{\bar{k}} J(v_i^{\bar{k}}, s_i^{\bar{k}}). \quad (6.14)$$

Compared to (5.24) we approximate the minimum of (6.2) by the energy of the last iterate $(v^{\bar{k}}, \sigma^{\bar{k}})$. Moreover, we replace the infinite infimal convolutions $R(v^k, \sigma^k)$ and $R(v^{\bar{k}}, \sigma^{\bar{k}})$ by the upper bounds $\sum_{i=1}^{N_k} c_i^k J(v_i^k, s_i^k)$ and $\sum_{i=1}^{N_{\bar{k}}} c_i^{\bar{k}} J(v_i^{\bar{k}}, s_i^{\bar{k}})$. This is justified, since [13, Theorem 4.4] guarantees that $\frac{1}{2} \|v^k - f\|_Y^2 + \alpha \sum_{i=1}^{N_k} c_i^k J(v_i^k, s_i^k)$ well approximates (6.2) if k is big enough.

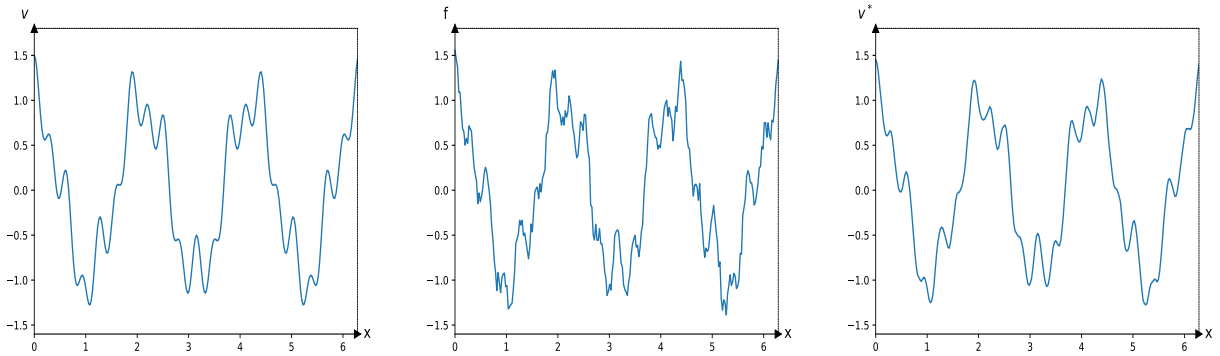


Figure 1: From left to right: the ground truth v ; the data f , i.e., the ground truth v corrupted by Gaussian noise; the reconstruction v^* provided by Algorithm 3.

Note that the obtained reconstruction resembles a Laplacian regularization of a noisy signal and it is hard to give a clear interpretation on the optimal orders $s \in [0, 1]$ selected by our algorithm. We stress however that the goal of this first example is solely to illustrate the main features of our general algorithm and demonstrate numerically the expected sublinear convergence of the residuals. We also note that in this example the rate of convergence appears to be linear, at least asymptotically. This has to be expected [13], but the theoretical verification of this behaviour is outside of the scope of this work.

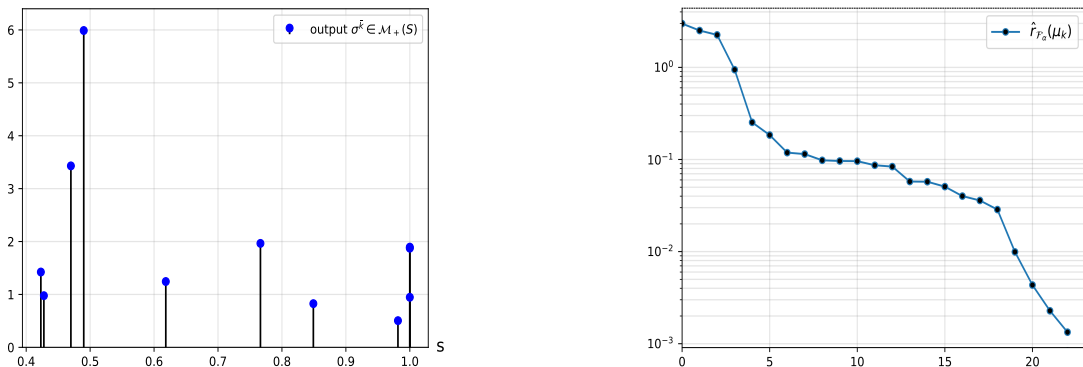


Figure 2: On the left: the measure $\sigma^{\bar{k}} \in \mathcal{M}_+(S)$ obtained in output by Algorithm 3; its support consists of the orders $s \in [0, 1]$ selected by Algorithm 3. On the right: a convergence graph, showing the approximate residual $\hat{r}_{\mathcal{F}_\alpha}$ (6.14) computed at each iteration.

6.2 Learning the anisotropies of fractional Laplacians in an infinite infimal convolution

In this section we design an infinite infimal convolution-type regularizer that is based on a variant of the anisotropic fractional Laplacian. The anisotropy is determined by a parameter $s \in [0, \pi]$ providing the direction $(\cos(s), \sin(s)) \in \mathbb{S}^1$. We construct an infinite infimal convolution regularizer that automatically selects the directions of higher oscillation of the noisy image f and applies a lower regularization in such directions. In this way, the regularizer learns how to preserve the anisotropy of the data and removes the noise that is assumed to be isotropic.

6.2.1 Setting

We denote by T^2 the 2-dimensional torus and consider $S = [0, \pi]$ and $X = L^2(T^2)$ endowed with the L^2 -norm. We define the operator $J : L^2(T^2) \times [0, \pi] \rightarrow [0, +\infty]$ as follows

$$J(v, s) := \sqrt{\sum_{m \in \mathbb{Z}^2} (|m \cdot (\cos(s), \sin(s))| + \zeta|m| + \omega)^{4\gamma} |\hat{v}(m)|^2}, \quad (6.15)$$

where $\gamma \in [0, 1]$, $\zeta \geq 0$ and $\omega > 0$ are fixed parameters.

Remark 6.4. The parameter $\omega > 0$ in the definition of J is essential for ensuring the coercivity of J at frequencies $m \in \mathbb{Z}^2$ such that $m \cdot (\cos(s), \sin(s)) = 0$. Since the penalization $\sqrt{\sum_{m \in \mathbb{Z}^2} (\zeta|m|)^{4\gamma} |\hat{v}(m)|^2}$ for $\zeta > 0$ models isotropic regularization of a specific rate, the corresponding term in J can therefore be useful to denoise images that exhibit isotropic behavior in some regions. If ζ is large then our regularizer acts similarly to a standard fractional Laplacian of order γ , since the anisotropic part is negligible. Further variants of (6.15) are possible. For example, one could allow for different orders of smoothness in a direction and its orthogonal, by considering the penalization

$$\sqrt{\sum_{m \in \mathbb{Z}^2} (|m \cdot (\cos(s), \sin(s))|^{4\gamma_1} + |m \cdot (-\sin(s), \cos(s))|^{4\gamma_2} + \omega) |\hat{v}(m)|^2}$$

for $\gamma_1 > \gamma_2 > 0$.

We consider the following regularizer given by the infinite infimal convolution of $J(\cdot, s)$:

$$R(v, \sigma) = \inf_{u \in L^1((S, \sigma); L^2(T^2))} \int_0^\pi J(u(s), s) d\sigma(s) + \mathbb{1}_{\{v\}}(\mathcal{I}_\sigma u), \quad (6.16)$$

and the corresponding denoising problem

$$\inf_{\substack{v \in L^2(T^2), \\ \sigma \in \mathcal{M}_+(S)}} \frac{1}{2} \|v - f\|_{L^2(T^2)}^2 + \alpha R(v, \sigma), \quad (6.17)$$

where $f \in L^2(T^2)$ is a given measurement.

We verify that this problem satisfies the assumptions made in Section 2.2.

Lemma 6.5. *The functional J defined in (6.15) satisfies (H1), (H2) and (H3) in Section 2.2.*

Proof. Assumption (H1) is straightforward from the definition of J . Assumption (H2) can be proven similarly to Lemma 6.3 using Fatou's lemma. Assumption (H3) follows immediately from the definition of J since $\omega > 0$. \square

6.2.2 The generalized conditional gradient method

We now describe a generalized conditional gradient method to solve (6.17), following the general procedure outlined in Section 5. For $B_s = \{v \in L^2(T^2) : \alpha J(v, s) \leq 1\}$, Lemma B.7 with the choice $f(m) = \alpha^2(|m \cdot (\cos(s), \sin(s))| + \zeta|m| + \omega)^{4\gamma}$ provides the following characterization of the extremal points of B_s :

$$\text{Ext}(B_s) = \left\{ v \in L^2(T^2) : \alpha \sqrt{\sum_{m \in \mathbb{Z}^2} (|m \cdot (\cos(s), \sin(s))| + \zeta|m| + \omega)^{4\gamma} |\hat{v}(m)|^2} = 1 \right\}, \quad (6.18)$$

where, in particular, we observe that $0 \notin \text{Ext}(B_s)$ for $s \in [0, \pi]$, cf. Remark 5.4. Therefore, given an iterate $\mu^k = \sum_{i=1}^{N_k} c_i^k v_i^k \delta_{s_i^k}$ with $N_k \in \mathbb{N}$, $c_i^k > 0$, $s_i^k \in [0, \pi]$ and $v_i^k \in \text{Ext}(B_s)$ as in (6.18), the insertion step in (5.18) becomes

$$\arg \max_{\substack{s \in [0, \pi], \\ v \in L^2(T^2)}} \left\langle f - \sum_{i=1}^{N_k} c_i^k v_i^k, v \right\rangle_{L^2(T^2)} + \mathbb{1}_{\{\alpha \sqrt{\sum_{m \in \mathbb{Z}^2} (|m \cdot (\cos(s), \sin(s))| + \zeta|m| + \omega)^{4\gamma} |\hat{v}(m)|^2} = 1\}}(v). \quad (6.19)$$

The following lemma gives an explicit formula for the inserted $v \in H^{2\gamma}(T^2)$ for a fixed $s \in [0, \pi]$.

Lemma 6.6. *Given $w \in L^2(T^2)$ and $s \in [0, \pi]$ consider the variational problem*

$$\max_{v \in L^2(T^2)} \langle w, v \rangle_{L^2(T^2)} + \mathbb{1}_{\{\alpha \sqrt{\sum_{m \in \mathbb{Z}^2} (|m \cdot (\cos(s), \sin(s))| + \zeta|m| + \omega)^{4\gamma} |\hat{v}(m)|^2} = 1\}}(v). \quad (6.20)$$

A solution of (6.20) can be computed as

$$v^*(x) = \frac{1}{A} \sum_{m \in \mathbb{Z}^2} (|m \cdot (\cos(s), \sin(s))| + \zeta|m| + \omega)^{-4\gamma} \hat{w}(m) e^{im \cdot x} \quad (6.21)$$

where

$$A = \alpha \sqrt{\sum_{m \in \mathbb{Z}^2} (|m \cdot (\cos(s), \sin(s))| + \zeta|m| + \omega)^{-4\gamma} |\hat{w}(m)|^2}. \quad (6.22)$$

Proof. Since the proof is similar to that of Lemma 6.3 we omit it. \square

In Algorithm 4 we rewrite the generalized conditional gradient method specialized to (6.17). We initialize the algorithm with a tuple of pairs in $X \times S$ denoted by $\Xi^0 = \{(v_1^0, s_1^0), \dots, (v_{N_0}^0, s_{N_0}^0)\}$ where $N_0 \in \mathbb{N}$, $s_i^0 \in S$ and $v_i^0 \in \text{Ext}(B_{s_i^0})$, and a tuple of positive weights $(c_1^0, \dots, c_{N_0}^0)$. In the insertion step at the k -th iteration of the algorithm, the newly inserted atom is computed by first applying Lemma 6.6 to find the optimal solution for a fixed $s \in [0, \pi]$, denoted by $v_{N_k+1}^k(s)$. Then, similarly to Section 6.1, $v_{N_k+1}^k(s)$ is optimized with respect to s by solving

$$\arg \max_{s \in [0, \pi]} \langle p^k, v_{N_k+1}^k(s) \rangle. \quad (6.23)$$

Similarly to Section 6.1, we remark that the computation of (6.23) is performed using a Python implementation of a basin-hopping-type algorithm whose performance is good enough to compute $\arg \max_{s \in [0, \pi]} \langle p^k, v_{N_k+1}^k(s) \rangle$ efficiently and with satisfactory accuracy.

6.2.3 Numerical experiments

In this section we present several numerical examples of the applications of Algorithm 4 for denoising images with high frequencies in certain directions. We show that the regularizer is able to learn such directions and preserve the related frequencies in the reconstruction. For each example we plot the output measure $\sigma^{\bar{k}} \in \mathcal{M}_+(S)$ that is supported on the $s \in [0, \pi]$ selected by our algorithm.

Algorithm 4 Generalized conditional gradient method for infinite infimal convolution of fractional Laplacians with adaptive anisotropies

Initialize: $v^0 = \sum_{i=1}^{N_0} c_i^0 v_i^0$, $\sigma^0 = \sum_{s \in \{s_i^0: i=1, \dots, N_0\}} \left\| \sum_{j \in \{1, \dots, N_0: s_j^0=s\}} c_j^0 v_j^0 \right\|_X \delta_s$
 $\Xi^0 = \{(v_1^0, s_1^0), \dots, (v_{N_0}^0, s_{N_0}^0)\}$ with (v_i^0, s_i^0) pairwise distinct, $c_i^0 > 0$, $s_i^0 \in S$, $v_i^0 \in \text{Ext}(B_{s_i^0})$ for $i = 1, \dots, N_0$.
for $k = 0, 1, 2 \dots$ **do**
 $p^k \leftarrow f - \sum_{i=1}^{N_k} c_i^k v_i^k$
 $v_{N_{k+1}}^k(s)(x) = \frac{\alpha^{-1}}{\sqrt{\sum_{m \in \mathbb{Z}^2} \frac{|\hat{p}^k(m)|^2}{(|m \cdot (\cos(s), \sin(s))| + \zeta|m| + \omega)^{4\gamma}}}} \sum_{m \in \mathbb{Z}^2} \frac{\hat{p}^k(m)}{(|m \cdot (\cos(s), \sin(s))| + \zeta|m| + \omega)^{4\gamma}} e^{im \cdot x}$
 $s_{N_{k+1}}^k \in \arg \max_{s \in [0, \pi]} \langle p^k, v_{N_{k+1}}^k(s) \rangle$
 $v_{N_{k+1}}^k \leftarrow v_{N_{k+1}}^k(s_{N_{k+1}}^k)$
if $k \geq 1$, $\langle p^k, v_{N_{k+1}}^k \rangle \leq 1$ or $(v_{N_{k+1}}^k, s_{N_{k+1}}^k) \in \Xi^k$ **then**
 return (v^k, σ^k)
end if
 $(\hat{c}_1, \dots, \hat{c}_{N_{k+1}}) \in \arg \min_{c \in \mathbb{R}_+^{N_{k+1}}} \frac{1}{2} \left\| \sum_{i=1}^{N_{k+1}} c_i v_i^k - f \right\|_{L^2(T^2)}^2 + \alpha \sum_{i=1}^{N_{k+1}} c_i$
 $v^{k+1} \leftarrow \sum_{i=1}^{N_{k+1}} \hat{c}_i v_i^k$, $\sigma^{k+1} \leftarrow \sum_{s \in \{s_i^k: i=1, \dots, N_{k+1}\}} \left\| \sum_{j \in \{1, \dots, N_{k+1}: s_j^k=s\}} \hat{c}_j v_j^k \right\|_X \delta_s$
 $\Xi^{k+1} \leftarrow \{(v_i^k, s_i^k) \in \Xi^k \cup \{(v_{N_{k+1}}^k, s_{N_{k+1}}^k)\} : \hat{c}_i^k > 0\}$
 $N_{k+1} \leftarrow \#\Xi^{k+1}$, $k \leftarrow k + 1$
end for
return (v^k, σ^k)

Note that the points $s = 0$ and $s = \pi$ correspond to the same direction $(\cos(s), \sin(s))$, and thus they should be identified. Moreover, similarly to the numerical experiments in Section 6.1.3, we plot an approximation of the residual $r_{\mathcal{F}_\alpha}$ (5.24) at each iteration in logarithmic scale. More precisely, denoting by $(v^k, \sigma^k) = (\sum_{i=1}^{N_k} c_i^k v_i^k, \sum_{s \in \{s_i^k: i=1, \dots, N_k\}} \left\| \sum_{j \in \{1, \dots, N_k: s_j^k=s\}} c_j^k v_j^k \right\|_X \delta_s)$ the iterate generated by Algorithm 4 and by $(v^{\bar{k}}, \sigma^{\bar{k}})$ the output, we plot the approximate residual $\hat{r}_{\mathcal{F}_\alpha}(\mu_k)$ defined as in (6.14). We refer to the discussion in Section 6.1.3 as a justification for this choice. In all examples we observe a convergence rate that is at least sublinear, as ensured by Theorem 5.7.

In the first example we aim to denoise a black-and-white grid of size 128×128 that is corrupted with Gaussian noise with standard deviation 0.3, see Figure 3. In the definition of J in (6.15) we set $\alpha = 5.5$, $\gamma = 0.25$, $\omega = 10^{-3}$, and $\zeta = 10^{-3}$. We compare the reconstruction obtained by our algorithm with a reconstruction produced using as regularizer the L^2 -norm of the $\frac{1}{4}$ -fractional Laplacian. We observe that our method reconstructs sharper edges. This is because our functional reduces the amount of regularization in the directions of the anisotropies of the image.

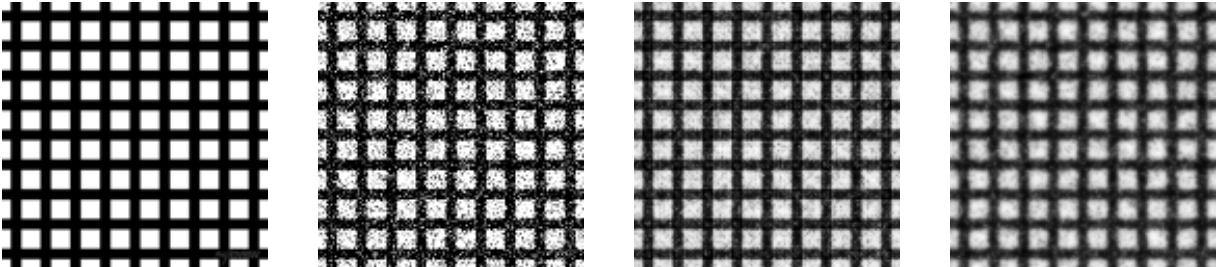


Figure 3: From left to right: ground truth; data f , i.e., ground truth corrupted by Gaussian noise; reconstruction obtained with our algorithm; reconstruction obtained with $\frac{1}{4}$ -fractional Laplacian regularization.

In Figure 4 we show the output measure $\sigma^{\bar{k}} \in \mathcal{M}_+(S)$ supported on the angles $s \in [0, \pi]$ selected by our algorithm and we plot the decay of the approximate residual $\hat{r}_{\mathcal{F}_\alpha}$ according to (6.14). We see that the selected angles concentrate around 0 , $\frac{\pi}{2}$ and π which are the most evident anisotropies of the image. Note again that the points $s = 0$ and $s = \pi$ correspond to the same direction $(\cos(s), \sin(s))$,

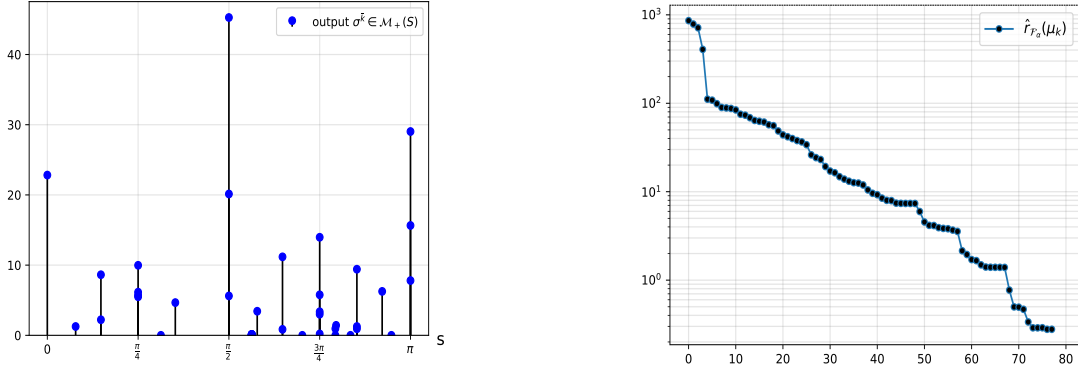


Figure 4: On the left: the measure $\sigma^{\bar{k}} \in \mathcal{M}_+(S)$ obtained in output by Algorithm 4; its support consists of the directions of the dominant anisotropies of the image $s \in [0, \pi]$ selected by Algorithm 4. On the right: a convergence graph, showing the approximate residual $\hat{r}_{\mathcal{F}_\alpha}$ according to (6.14) computed at each iteration.

and thus they should be identified. We also observe that due to the imperfect anisotropy of the grid, the optimal $\sigma \in \mathcal{M}_+(S)$ also has significant mass at $\frac{\pi}{4}$ and $\frac{3}{4}\pi$. The convergence plot illustrates the sublinear convergence of our algorithm.

In the second example we denoise a more complicated black-and-white geometric pattern corrupted by Gaussian noise with standard deviation equal to 0.3, see Figure 5. In the definition of J in (6.15) we set $\alpha = 5.5$, $\gamma = 0.25$, $\omega = 10^{-3}$, and $\zeta = 10^{-3}$. Similarly to the first example, we observe that our method reconstructs sharper edges thanks to its adaptivity. In Figure 5 we show the directions selected by the algorithm and the convergence plot of the approximate residual (5.24). In this example the selected $s \in [0, \pi]$ cluster around the values $\frac{1}{4}\pi$ and $\frac{3}{4}\pi$ that represent the anisotropies of the original image. The approximate convergence rate $\hat{r}_{\mathcal{F}_\alpha}$ appears again sublinear.

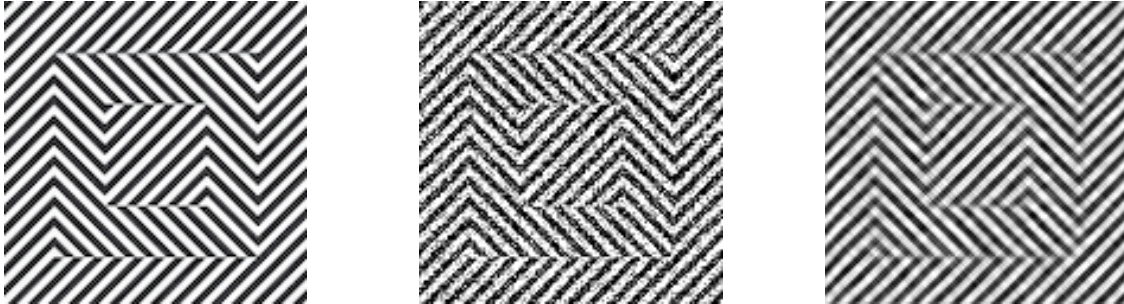


Figure 5: From left to right: the ground truth; the data f , i.e., the ground truth corrupted by Gaussian noise; the reconstruction obtained with our algorithm.

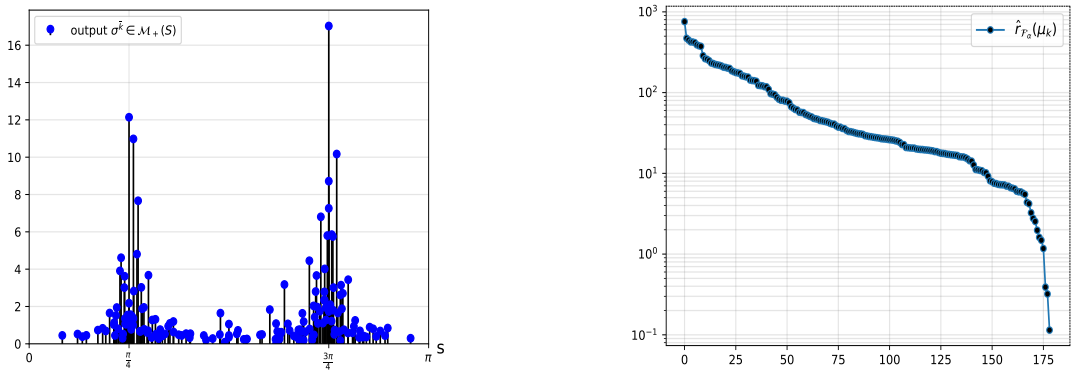


Figure 6: On the left: the measure $\sigma^{\bar{k}} \in \mathcal{M}_+(S)$ obtained in output by Algorithm 4; its support consists of the directions of the dominant anisotropies of the image $s \in [0, \pi]$ selected by Algorithm 4. On the right: a convergence graph, showing the approximate residual $\hat{r}_{\mathcal{F}_\alpha}$ according to (6.14) computed at each iteration.

In the last experiment we denoise a 253×253 grayscale image of a bamboo fence corrupted by Gaussian noise with standard deviation 0.2, see Figure 7. In the definition of J in (6.15) we set $\alpha = 6.5$, $\gamma = 0.25$, $\omega = 10^{-3}$, and $\zeta = 5 \cdot 10^{-3}$. Similarly to previous examples, we observe that our method reconstructs sharper edges. For this experiment we compare the reconstructions obtained by our method with classical model-based denoising methods. Firstly, we consider methods that do not impose preferred directions of regularization on the image: Total Variation (TV) [42] and Total Generalized Variation (TGV) [8], see Figure 8.

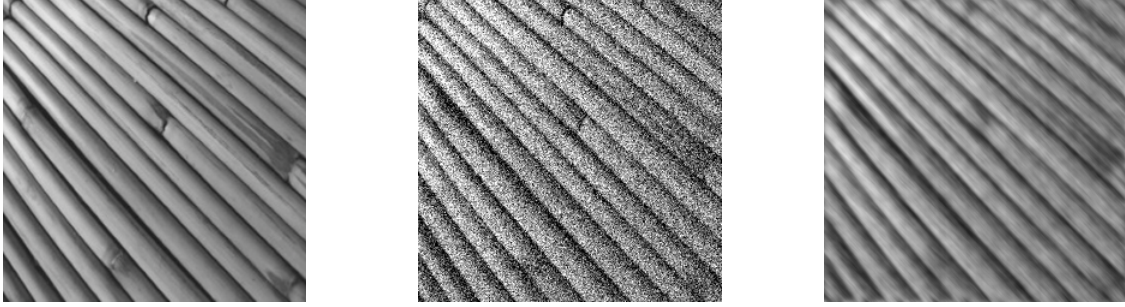


Figure 7: From left to right: ground truth; data f , i.e., ground truth corrupted by Gaussian noise with standard deviation 0.2 (PSNR = 14.0); reconstruction obtained with our algorithm (PSNR = 25.8).

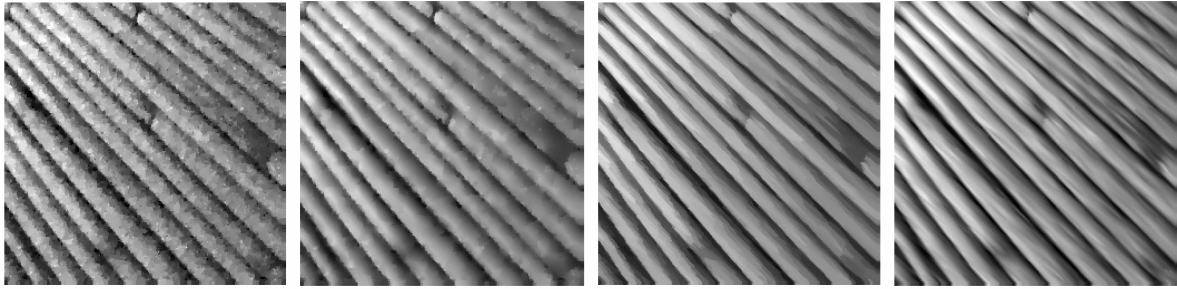


Figure 8: Comparison with classical denoising regularizers. From left to right: TV (PSNR = 23.8), TGV (PSNR = 24.7), Directional TV (PSNR = 26.8), and Directional TGV (PSNR = 28.2)

Our method outperforms them in terms of the PSNR and visual quality. Secondly, we compare our method to approaches that are aware of preferred directions in the image: directional TV [34] and directional TGV [34]. In this case, we achieve slightly worse performance in terms of PSNR and visual quality. This is due to the fact that TV and TGV regularization produce better

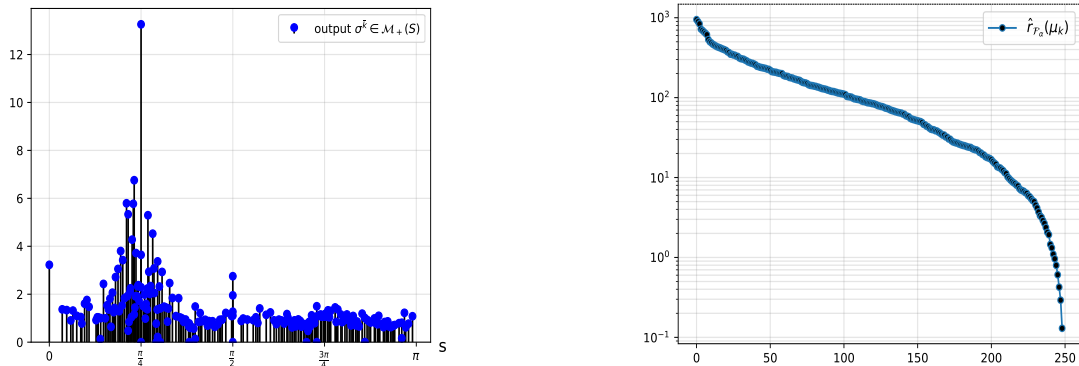


Figure 9: On the left: the measure $\sigma^{\bar{k}} \in \mathcal{M}_+(S)$ obtained in output by Algorithm 4; its support consists of the directions of the dominant anisotropies of the image $s \in [0, \pi]$ selected by Algorithm 4. On the right: a convergence graph, showing the approximate residual $\hat{r}_{\mathcal{F}_\alpha}$ according to (6.14) computed at each iteration.

reconstructions than the (isotropic) fractional Laplacian when the ground truths exhibit piecewise

constant behavior [5]. Therefore, for images like the bamboo fence, we cannot expect to outperform directional TV and directional TGV. However, we remark that, in contrast to [34] our method is able to select automatically the directions of the anisotropies of the image that are encoded into the output measure $\sigma \in \mathcal{M}_+(S)$. Figure 9 shows the directions selected by our algorithm and the convergence plot of the residual. As expected, the selected $s \in [0, \pi]$ concentrate around $\frac{\pi}{4}$ and we again observe a sublinear convergence rate for the residual.

7 Conclusions

We introduced the infinite infimal convolution operator for the regularization of ill-posed inverse problems. Such operator is constructed through an infimal convolution of a continuously parametrized family of convex, positively one-homogeneous and coercive functionals and the regularization depends on a Radon measure defined on the set of parameters S that prescribes which functionals are active in the infimal convolution. The variational problem is then formulated with respect to both the measure σ in S and the reconstruction is obtained by the integration operation $v = \int_S u(s) d\sigma(s)$. Using the positive one-homogeneity of each functional in the infimal convolution, we proposed a convex lifting of the problem in the space of measures with values in Banach space, which allows us to prove well-posedness. Moreover, since this reformulation leads to a convex optimization problem in the space of Radon measures, we could analyze the sparsity of solutions, obtain a representer theorem and propose a generalized conditional gradient method to compute the minimizers. This allows us to exploit sparsity both in the parameters space and in the reconstruction domain. Finally, we applied the model and the generalized conditional gradient method to several illustrative examples involving fractional-Laplacian-type regularizers.

The analysis of other types of regularizers (not of fractional-Laplacian type) is left to future work. Successful design of a generalized conditional gradient method relies on the possibility to solve the insertion step (5.18) in a fast and accurate way for a given $s \in S$. This is true for the fractional Laplacian (see Lemma 6.3 and Lemma 6.6); however, in case on more complicated regularizers, such as directional BV-norms and directional total generalized variations, the resolution of the insertion step could be impractical already for two-dimensional domains and high resolution images. Since there are no general recipes for solving (5.18) efficiently, the feasibility of Algorithm 2 and strategies of solving it need to be studied case-by-case.

8 Acknowledgments

KB gratefully acknowledges support by the Austrian Science Fund (FWF) through project P 29192 “Regularization graphs for variational imaging”. MC is supported by the Royal Society (Newton International Fellowship NIF\R1\192048 “Minimal partitions as a robustness boost for neural network classifiers”). The Institute of Mathematics and Scientific Computing, to which KB and MH are affiliated, is a member of NAWI Graz (<https://www.nawigraz.at/en/>). KB and MH are further members of BioTechMed Graz (<https://biotechmedgraz.at/en/>). YK acknowledges support of the EPSRC (Fellowship EP/V003615/1 and Programme Grant EP/V026259/1) and the National Physical Laboratory. CBS acknowledges support from the Philip Leverhulme Prize, the Royal Society Wolfson Fellowship, the EPSRC advanced career fellowship EP/V029428/1, EPSRC grants EP/S026045/1 and EP/T003553/1, EP/N014588/1, EP/T017961/1, the Wellcome Innovator Awards 215733/Z/19/Z and 221633/Z/20/Z, the European Union Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 777826 NoMADS, the Cantab Capital Institute for the Mathematics of Information, and the Alan Turing Institute. MC, YK, and CBS would like to thank the Isaac Newton Institute for Mathematical Sciences in Cambridge (supported by EPSRC grant no EP/R014604/1) for support and hospitality during the programme “Mathematics of Deep Learning” where part of the work on this paper was undertaken.

References

- [1] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Science Publications, 2000.
- [2] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*. Birkhäuser Basel, 2005.
- [3] H. Antil and S. Bartels. “Spectral approximation of fractional PDEs in image processing and phase field modeling”. In: *Computational Methods in Applied Mathematics* 17.4 (2017), pp. 661–678.
- [4] F. Balle et al. “Computation and visualization of local deformation for multiphase metallic materials by infimal convolution of TV-type functionals”. In: *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer, 2015, pp. 385–396.
- [5] S. Bartels and N. Weber. “Parameter learning and fractional differential operators: Applications in regularized image denoising and decomposition problems”. In: *Mathematical Control and Related Fields* 13.1 (2023), pp. 35–62.
- [6] F. Bartolucci, E. D. Vito, L. Rosasco, and S. Vigogna. “Understanding neural networks with reproducing kernel Banach spaces”. In: *Applied and Computational Harmonic Analysis* 62 (2023), pp. 194–236.
- [7] C. Boyer et al. “On representer theorems and convex regularization”. In: *SIAM Journal on Optimization* 29.2 (2019), pp. 1260–1281.
- [8] K. Bredies, K. Kunisch, and T. Pock. “Total generalized variation”. In: *SIAM Journal on Imaging Sciences* 3.3 (2010), pp. 492–526.
- [9] K. Bredies and H. Pikkarainen. “Inverse problems in spaces of measures”. In: *ESAIM: Control, Optimisation and Calculus of Variations* 19.1 (2013), pp. 190–218.
- [10] K. Bredies and M. Carioni. “Sparsity of solutions for variational inverse problems with finite-dimensional data”. In: *Calculus of Variations and Partial Differential Equations* 59.1 (2020), pp. 1–26.
- [11] K. Bredies, M. Carioni, and S. Fanzon. “A superposition principle for the inhomogeneous continuity equation with Hellinger–Kantorovich-regular coefficients”. In: *Communications in Partial Differential Equations* 47.10 (2022), pp. 2023–2069.
- [12] K. Bredies, M. Carioni, S. Fanzon, and F. Romero. “A generalized conditional gradient method for dynamic inverse problems with optimal transport regularization”. In: *Foundations of Computational Mathematics* (2022), pp. 1–66. <https://doi.org/10.1007/s10208-022-09561-z>.
- [13] K. Bredies, M. Carioni, S. Fanzon, and D. Walter. “Asymptotic linear convergence of fully-corrective generalized conditional gradient methods”. In: *arXiv preprint arXiv:2110.06756* (2023).
- [14] K. Bredies, M. Carioni, and M. Holler. “Regularization graphs — a unified framework for variational regularization of inverse problems”. In: *Inverse Problems* 38.10 (2022), p. 105006.
- [15] K. Bredies and D. A. Lorenz. “Iterated Hard Shrinkage for Minimization Problems with Sparsity Constraints”. In: *SIAM Journal on Scientific Computing* 30.2 (2008), pp. 657–683.
- [16] K. Bredies, D. A. Lorenz, and P. Maass. “A generalized conditional gradient method and its connection to an iterative shrinkage method”. In: *Computational Optimization and Applications* 42.2 (2009), pp. 173–193.
- [17] M. Burger, Y. Korolev, C.-B. Schönlieb, and C. Stollenwerk. “A total variation based regularizer promoting piecewise-Lipschitz reconstructions”. In: *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer, 2019, pp. 485–497.
- [18] M. Burger, K. Papafitsoros, E. Papoutsellis, and C.-B. Schönlieb. “Infimal Convolution Regularisation Functionals of BV and L^p Spaces. The Case $p = \infty$ ”. In: *IFIP Conference on System Modeling and Optimization*. Springer, 2015, pp. 169–179.
- [19] M. Burger, K. Papafitsoros, E. Papoutsellis, and C.-B. Schönlieb. “Infimal Convolution Regularisation Functionals of BV and L^p Spaces. Part I: The Finite p Case”. In: *J. Math. Imaging Vis.* 55 (2016), p. 343.
- [20] M. Carioni, J. A. Iglesias, and D. Walter. “Extremal points and sparse optimization for generalized Kantorovich-Rubinstein norms”. In: *arXiv preprint arXiv:2209.09167* (2022).
- [21] C. Castaing, P. R. De Fitte, and M. Valadier. *Young measures on topological spaces: with applications in control theory and probability theory*. Vol. 571. Springer Science & Business Media, 2004.

- [22] A. Chambolle and P.-L. Lions. “Image recovery via total variation minimization and related problems”. In: *Numerische Mathematik* 76.2 (1997), pp. 167–188.
- [23] T. F. Chan, S. Esedoglu, and F. Park. “A fourth order dual method for staircase reduction in texture extraction and image restoration problems”. In: *2010 IEEE International Conference on Image Processing*. IEEE. 2010, pp. 4137–4140.
- [24] E. Di Nezza, G. Palatucci, and E. Valdinoci. “Hitchhiker’s guide to the fractional Sobolev spaces”. In: *Bulletin des sciences mathématiques* 136.5 (2012), pp. 521–573.
- [25] J. Diestel and J. J. Uhl. *Vector Measures*. Mathematical Surveys. American Mathematical Society, 1977.
- [26] V. Duval and R. Tovey. “Dynamical Programming for off-the-grid dynamic Inverse Problems”. In: *arXiv preprint arXiv:2112.11378* (2021).
- [27] A. Flinth, F. De Gournay, and P. Weiss. “On the linear convergence rates of exchange and continuous methods for total variation minimization”. In: *Mathematical Programming* 190.1 (2021), pp. 221–257.
- [28] M. Frank and P. Wolfe. “An algorithm for quadratic programming”. In: *Naval research logistics quarterly* 3.1-2 (1956), pp. 95–110.
- [29] Y. Gao and K. Bredies. “Infimal convolution of oscillation total generalized variation for the recovery of images with structured texture”. In: *SIAM Journal on Imaging Sciences* 11.3 (2018), pp. 2021–2063.
- [30] Y. Gao and X. Yang. “Infimal convolution type regularization of TGV and shearlet transform for image restoration”. In: *Comput. Vis. Image Underst.* 182 (2019), pp. 38–49.
- [31] H. M. Hochman and J. D. Rodgers. “Pareto optimal redistribution”. In: *The American economic review* 59.4 (1969), pp. 542–557.
- [32] M. Holler and K. Kunisch. “On Infimal Convolution of TV-Type Functionals and Applications to Video and Image Reconstruction”. In: *SIAM Journal on Imaging Sciences* 7.4 (2014), pp. 2258–2300.
- [33] J. A. Iglesias and D. Walter. “Extremal points of total generalized variation balls in 1D: characterization and applications”. In: *arXiv preprint arXiv:2112.06846* (2021).
- [34] R. D. Kongskov, Y. Dong, and K. Knudsen. “Directional total generalized variation regularization”. In: *BIT Numerical Mathematics* 59.4 (2019), pp. 903–928.
- [35] G. Kutyniok. “Clustered sparsity and separation of cartoon and texture”. In: *SIAM J. Imaging Sci.* 6.2 (2013), pp. 848–874.
- [36] L. Meziani. “On the dual space $C_0^*(S, X)$ ”. In: *Acta Mathematica Universitatis Comenianae* 78 (2009), pp. 153–160.
- [37] S. Parisotto, J. Lellmann, S. Masnou, and C.-B. Schönlieb. “Higher-Order Total Directional Variation: Imaging Applications”. In: *SIAM Journal on Imaging Sciences* 13.4 (2020), pp. 2063–2104.
- [38] S. Parisotto, S. Masnou, and C.-B. Schönlieb. “Higher-Order Total Directional Variation: Analysis”. In: *SIAM Journal on Imaging Sciences* 13.1 (2020), pp. 474–496.
- [39] K. Pieper and D. Walter. “Linear convergence of accelerated conditional gradient algorithms in spaces of measures”. In: *ESAIM: Control, Optimisation and Calculus of Variations* 27 (2021), p. 38.
- [40] Y. G. Reshetnyak. “Weak convergence of completely additive vector functions on a set”. In: *Siberian Mathematical Journal* 9.6 (1968), pp. 1039–1045.
- [41] R. T. Rockafellar. *Conjugate Duality and Optimization*. Society for Industrial and Applied Mathematics, 1974.
- [42] L. I. Rudin, S. Osher, and E. Fatemi. “Nonlinear total variation based noise removal algorithms”. In: *Physica D: Nonlinear Phenomena* 60.1-4 (1992), pp. 259–268.
- [43] R. A. Ryan. *Introduction to tensor products of Banach spaces*. Springer Science & Business Media, 2013.
- [44] M. Schloegl et al. “Infimal convolution of total generalized variation functionals for dynamic MRI”. In: *Magnetic Resonance in Medicine* 78.1 (2017), pp. 142–155.
- [45] S. Setzer, G. Steidl, and T. Teuber. “Infimal convolution regularizations with discrete ℓ_1 -type functionals”. In: *Communications in Mathematical Sciences* 9.3 (2011), pp. 797–827.
- [46] I. Singer. “Linear functionals on the space of continuous mappings of a compact Hausdorff space into a Banach spaces”. In: *Rev. Math. Pures Appl.* 2 (1957), pp. 301–315.

- [47] J. E. Stiglitz. “Pareto efficient and optimal taxation and the new new welfare economics”. In: *Handbook of public economics*. Vol. 2. Elsevier, 1987, pp. 991–1042.
- [48] M. Unser. “A Unifying Representer Theorem for Inverse Problems and Machine Learning”. In: *Foundations of Computational Mathematics* 21.4 (2021), pp. 941–960.
- [49] M. Unser, J. Fageot, and J. P. Ward. “Splines are universal solutions of linear inverse problems with generalized TV regularization”. In: *SIAM Review* 59.4 (2017), pp. 769–793.
- [50] P. G. Warr. “Pareto optimal redistribution and private charity”. In: *Journal of Public Economics* 19.1 (1982), pp. 131–138.

A Measures with values in Banach spaces

In this section we give the necessary definitions and results about measures with values in Banach spaces that are needed in this paper and we formulate them in our setting. We refer to [25, 43] to a more detailed overview about the existing theory.

A.1 L^p -functions with values in Banach spaces

Let $d \in \mathbb{N}$ and X be a Banach space. Given a non-empty compact set $S \subset \mathbb{R}^d$, $\mu \in \mathcal{M}_+(S)$ a Radon measure on S and $p \in [1, \infty)$ we denote by $L^p((S, \mu); X)$ the space of L^p functions with values in X defined as

$$L^p((S, \mu); X) = \left\{ u : S \rightarrow X \text{ measurable} : \int_S \|u(s)\|_X^p d\mu(s) < \infty \right\}, \quad (\text{A.1})$$

where we say that a function $u : S \rightarrow X$ is measurable if it is the μ -a.e. limit of a sequence of simple functions; see [43, Section 2.3] for more details. Note also that each $u \in L^p((S, \mu); X)$ is integrable in the sense of Bochner. It is well-known that the space $L^p((S, \mu); X)$ is a Banach space with the norm

$$\|u\|_{L^p((S, \mu); X)} = \left(\int_S \|u(s)\|_X^p d\mu(s) \right)^{1/p}.$$

A.2 Definition and Radon–Nikodým property

We will consider finite Radon measures defined in the compact set $S \subset \mathbb{R}^d$ with values in X that is the dual of a separable Banach space X_* and we denote this set as $\mathcal{M}(S; X)$. The formal definition is similar to the classical notion of measure with values in \mathbb{R} . Denoting by Σ the Borel σ -algebra of S a function $\mu : \Sigma \rightarrow X$ is a Borel measure if

i) $\mu(\emptyset) = 0$

ii) For any countable family $(E_i)_{i=1}^\infty$, of pairwise disjoint Borel sets we have that

$$\mu \left(\bigcup_{i=1}^\infty E_i \right) = \sum_{i=1}^\infty \mu(E_i), \quad (\text{A.2})$$

where the limit in (A.2) is taken with respect to the strong topology of X .

The classical definition of variation can be easily extended to measures with values in Banach spaces. Given a Borel measure μ and $E \in \Sigma$ we define

$$|\mu|(E) = \sup \left\{ \sum_{i=1}^n \|\mu(A_i)\|_X : n \in \mathbb{N}, (A_i)_{i=1}^n \text{ partition of } E \right\}. \quad (\text{A.3})$$

We denote by $\mathcal{M}(S; X)$ the set of Borel measures with finite variation and we call their elements Radon measures. It can be shown that for $\mu \in \mathcal{M}(S; X)$, $|\mu|$ belongs to the set of finite positive

Borel measures $\mathcal{M}_+(S)$ [25, Proposition 9] and $\mathcal{M}(S; X)$ is a Banach space with the total variation norm defined as $\|\mu\|_{\mathcal{M}(S; X)} = |\mu|(S)$ for every $\mu \in \mathcal{M}(S; X)$. We remark that every $\mu \in \mathcal{M}(S; X)$ is regular, in the sense that $|\mu|$ is regular in the classical sense, being a finite positive Borel measure defined on a compact set in \mathbb{R}^d . Moreover, for every $\mathcal{M}(S; X)$ we have $\mu \ll |\mu|$ in the sense that for every Borel set $E \subset S$ such that $|\mu|(E) = 0$ it holds that $\mu(E) = 0$.

Remark A.1. We point out that the total variation norm defined above can be replaced by the total semi-variation [25] to endow the space $\mathcal{M}(S; X)$ with an alternative Banach space structure. As we consider the integral of strongly measurable functions with respect to measures in $\mathcal{M}(S; X)$ the natural norm to be considered is the total variation norm. We refer to [25, 43] for the definition of the total semi-variation and the corresponding theory of weak integration.

In the next proposition we deal with vector measures constructed as

$$\mu(E) = \int_E u(s) d\sigma(s), \quad (\text{A.4})$$

for all $E \subset S$ Borel, where $\sigma \in \mathcal{M}_+(S)$ and $u \in L^1((S, \sigma); X)$. Note that the integral above is defined as the Bochner integral of $u \in L^1((S, \sigma); X)$. Following the notation for classical measures, given $\mu \in \mathcal{M}(S; X)$ defined as in (A.4) we write $\mu = u\sigma$. Moreover, we call u the density of μ with respect to ν and we denote it by $u = \frac{\mu}{\sigma}$. Note that u is unique σ almost everywhere due to [25, Corollary 5].

Proposition A.2. *Let Σ be the Borel σ -algebra of S , $\sigma \in \mathcal{M}_+(S)$ and $u \in L^1((S, \sigma); X)$. Then, the map $\mu : \Sigma \rightarrow X$ defined in (A.4) belongs to $\mathcal{M}(S; X)$ and its variation is given by*

$$|\mu|(E) = \int_E \|u(s)\|_X d\sigma(s) \quad (\text{A.5})$$

for every $E \subset S$ measurable set. In particular, $\|u(s)\|_X > 0$ for $|\mu|$ -a.e. $s \in S$. Moreover, $\mu \ll |\mu|$ and for every $E \subset S$ measurable,

$$\mu(E) = \int_E \hat{u}(s) d|\mu|(s), \quad (\text{A.6})$$

where $\hat{u} \in L^1((S, |\mu|); X)$ is defined as $\hat{u}(s) = \frac{u(s)}{\|u(s)\|_X}$ for $|\mu|$ -a.e. $s \in S$.

Proof. First the verification that μ belongs to $\mathcal{M}(S; X)$ is straightforward and (A.5) follows directly from Proposition 5.4 in [43]. Notice that by the properties of the Bochner integral we have the following inequality

$$\|\mu(E)\|_X \leq \int_E \|u(s)\|_X d\sigma(s) = |\mu|(E) \quad (\text{A.7})$$

implying that $\mu \ll |\mu|$. We now show (A.6). Indeed, thanks to (A.5) we first notice that $\|u(s)\|_X > 0$ for $|\mu|$ -a.e. $s \in S$. Thus, for every measurable set E there holds

$$\int_E \frac{u(s)}{\|u(s)\|_X} d|\mu|(s) = \int_E u(s) d\sigma(s) = \mu(E),$$

as we wanted to prove. \square

An important property for measures with values in a Banach space X is the validity of the Radon–Nikodým theorem.

Definition A.3 (Radon–Nikodým property). The Banach space X has the Radon–Nikodým property if for every set Ω , every σ -algebra Σ on Ω and every measure $\mu : \Sigma \rightarrow \mathbb{R}$ of bounded variation such that $\mu \ll \sigma$ for $\sigma : \Sigma \rightarrow \mathbb{R}$ a finite positive measure, there exists $u \in L^1((\Omega, \sigma); X)$ such that

$$\mu(E) = \int_E u(s) d\sigma(s) \quad \text{for every } E \in \Sigma. \quad (\text{A.8})$$

The Radon–Nikodým property is true for separable dual Banach spaces due to the classical Dunford–Pettis theorem, see Theorem 1 [25, p. III.3].

Theorem A.4 (Dunford–Pettis). *Separable dual Banach spaces have the Radon–Nikodým property.*

We conclude this section with the Lebesgue decomposition theorem adapted to measures with values in Banach spaces [25, Theorem 9].

Theorem A.5 (Lebesgue decomposition theorem). *Given $\mu \in \mathcal{M}(S; X)$ and $\sigma \in \mathcal{M}_+(S)$, there exist unique $\mu^a, \mu^s \in \mathcal{M}(S; X)$ such that*

- $\mu^a \ll \sigma$,
- $\varphi \circ \mu^s \in \mathcal{M}(S)$ and σ are mutually singular for every $\varphi \in X^*$,
- $\mu = \mu^a + \mu^s$.

Additionally, $|\mu| = |\mu^a| + |\mu^s|$, and $|\mu^s|$ and σ are mutually singular.

A.3 Duality theory

Define the space $C(S; X)$ to be the set of continuous functions on S with values in X endowed with the norm topology given by

$$\|T\|_{C(S; X)} := \sup_{s \in S} \|T(s)\|_X \quad (\text{A.9})$$

for $T \in C(S; X)$. Here we are interested in the identification of $C(S; X)^*$ with the space of vector measures $\mathcal{M}(S; X^*)$, if S is a non-empty compact set of \mathbb{R}^d . This is a classical result due to Singer [46] that we state here for the reader's convenience. See also, for instance, [36] or [25, Section 6.5]. Note that since S is non-empty and compact, the space $\mathcal{M}(S; X^*)$ coincides with the space of regular X^* -valued Borel measures of finite total variation.

Theorem A.6. *Let S be a compact Hausdorff space and X a Banach space. Then, there exists an isometric isomorphism T between $\mathcal{M}(S; X^*)$ and $C(S; X)^*$ such that*

$$T(\mu)(f) = \int_S f d\mu, \quad \text{for every } \mu \in \mathcal{M}(S, X^*), f \in C(S; X)$$

and

$$\|\mu\|_{\mathcal{M}(S, X^*)} = \|T(\mu)\|_{C(S, X)^*} = \sup_{\|f\|_{C(S; X)} \leq 1} \int_S f d\mu, \quad \text{for every } \mu \in \mathcal{M}(S, X^*).$$

In light of Theorem A.6 we say that a sequence of regular X^* -valued vector measures on S denoted by $\{\mu^n\}_n$ weakly* converges to μ if

$$\lim_{n \rightarrow +\infty} \int_S f d\mu^n = \int_S f d\mu \quad \text{for every } f \in C(S; X). \quad (\text{A.10})$$

We also have the following remark.

Remark A.7. Theorem A.6 and the definition of weak* convergence in (A.10) imply that the total variation norm of a regular X^* -valued Borel measure on S is weakly-* lower semicontinuous. Moreover choosing $f(s) = v \in X$ in (A.10) we obtain that

$$\lim_{n \rightarrow +\infty} \langle \mu^n(S), v \rangle = \lim_{n \rightarrow +\infty} \int_S v d\mu^n = \int_S v d\mu = \langle \mu(S), v \rangle$$

for every $v \in X$, implying that $\mu^n(S)$ converges weakly* to $\mu(S)$.

B Fractional-Laplacian-type operators

In this section of the appendix we recall basic results and definitions about fractional Sobolev spaces, fractional Laplacian operators as well as necessary functional analytic tools required in the paper. We also consider the energies obtained as the L^2 -norm of fractional-Laplacian-type and characterize the extremal points of their balls. For more details about fractional Sobolev spaces and fractional Laplacians we refer to [24].

B.1 Periodic L^2 -functions

For $q \in \mathbb{N}$ we denote by T^q the (2π) -periodic torus defined by a quotient procedure as $(\mathbb{R}/2\pi\mathbb{Z})^q$. We denote by $L^2(T^q)$ the set of L^2 functions on T^q defined as

$$L^2(T^q) = \left\{ v : T^q \rightarrow \mathbb{R} \text{ measurable} : \int_{T^q} |v(x)|^2 dx < \infty \right\}. \quad (\text{B.1})$$

We identify the space $L^2(T^q)$ with the space of equivalence classes of $[0, 2\pi)^q$ -periodic functions $v : \mathbb{R}^q \rightarrow \mathbb{R}$ such that

$$\int_{(0, 2\pi)^q} |v(x)|^2 dx < \infty. \quad (\text{B.2})$$

We endow $L^2(T^q)$ with a Hilbert structure by means of the scalar product

$$\langle v, w \rangle_{L^2(T^q)} = \int_{(0, 2\pi)^q} v(x)w(x) dx \quad \text{for } v, w \in L^2(T^q) \quad (\text{B.3})$$

inducing the norm

$$\|v\|_{L^2(T^q)} = \sqrt{\int_{(0, 2\pi)^q} |v(x)|^2 dx}. \quad (\text{B.4})$$

The space $L^2(T^q)$ is a separable Hilbert space and its predual is $L^2(T^q)$ itself. We also introduce the L^2 -periodic functions with zero mean, defined as

$$L^2_{\circ}(T^q) = \left\{ u \in L^2(T^q) : \int_{(0, 2\pi)^q} u(x) dx = 0 \right\}. \quad (\text{B.5})$$

Notice that $L^2_{\circ}(T^q)$ is a closed subspace of $L^2(T^q)$. Therefore, $L^2_{\circ}(T^q)$ is a separable Hilbert space as well.

B.2 Fourier series of periodic L^2 -functions

For $m \in \mathbb{Z}^q$ we define the m -th Fourier coefficient of $v \in L^2(T^q)$ as

$$\mathcal{F}v(m) = \hat{v}(m) = \frac{1}{(2\pi)^q} \int_{(0, 2\pi)^q} v(x) e^{-ix \cdot m} dx. \quad (\text{B.6})$$

It is standard to show that for every $v \in L^2(T^q)$ the Fourier series $\sum_{m \in \mathbb{Z}^q} \hat{v}(m) e^{ix \cdot m}$ converges to v in $L^2(T^q)$ and we write $v(x) = \sum_{m \in \mathbb{Z}^q} \hat{v}(m) e^{ix \cdot m}$. Moreover, Parseval's theorem holds for every $v, w \in L^2(T^q)$, i.e.,

$$\langle v, w \rangle_{L^2(T^q)} = (2\pi)^q \langle \hat{v}, \hat{w} \rangle_{\ell^2(\mathbb{Z}^q)}. \quad (\text{B.7})$$

B.3 Periodic fractional Sobolev functions

We now define fractional Sobolev functions on the periodic domain T^q for $q \in \mathbb{N}$.

Definition B.1 (Fractional Sobolev functions on periodic domains). Given a function $v : T^q \rightarrow \mathbb{R}$ and $s \geq 0$, we say that $v \in H^s(T^q)$ if $v \in L^2(T^q)$ and

$$[v]_{H^s(T^q)}^2 := \sum_{m \in \mathbb{Z}^q} |m|^{2s} |\hat{v}(m)|^2 < \infty. \quad (\text{B.8})$$

The space $H^s(T^q)$ is a Hilbert space with the scalar product defined as

$$\langle v, w \rangle_{H^s(T^q)} = \langle v, w \rangle_{L^2(T^q)} + \sum_{m \in \mathbb{Z}^q} |m|^{2s} \hat{v}(m) \overline{\hat{w}(m)} \quad (\text{B.9})$$

inducing a norm defined as $\|v\|_{H^s(T^q)} = \sqrt{\|v\|_{L^2(T^q)}^2 + [v]_{H^s(T^q)}^2}$.

Remark B.2. Notice that for $s = 0$ the space $H^s(T^q)$ is isomorphic to $L^2(T^q)$ and $\|v\|_{H^0(T^q)} = \sqrt{1 + (2\pi)^{-q}} \|v\|_{L^2(T^q)}$ for every $v \in L^2(T^q)$.

B.4 The fractional Laplacian

We define the fractional Laplacian in periodic domains as follows.

Definition B.3 (Fractional Laplacian). Given $v \in H^{2s}(T^q)$ and $s \in (0, 1]$ we define the fractional Laplacian of v as

$$(-\Delta^s)v(x) = \sum_{m \in \mathbb{Z}^q} |m|^{2s} \hat{v}(m) e^{ix \cdot m}. \quad (\text{B.10})$$

Remark B.4. According to the previous definitions it is easy to check that $(-\Delta^s)v \in L^2(T^q)$ with

$$\|(-\Delta^s)v\|_{L^2(T^q)} = (2\pi)^{q/2} [v]_{H^{2s}(T^q)}. \quad (\text{B.11})$$

Using Parseval's theorem it is easy to show that the following version of the Poincaré inequality holds.

Theorem B.5 (Poincaré inequality). *For every $0 \leq s \leq 1$ there holds*

$$\|v\|_{L^2(T^q)} \leq \|(-\Delta^s)v\|_{L^2(T^q)} \quad (\text{B.12})$$

for every $v \in H^{2s}(T^q)$ with zero mean.

B.5 Extremal points of L^2 -balls of fractional-Laplacian-type operators

In this paper we will consider L^2 -norms of fractional-Laplacian-type operators. This section is devoted to characterize the extremal points of their L^2 balls. Let $f : \mathbb{Z}^q \rightarrow [0, \infty)$ be a function such that

- i) For every $m \neq 0$ it holds that $f(m) > 0$.

We consider the convex, positively one-homogeneous functional $J_f : L^2(T^q) \rightarrow [0, \infty)$ defined as

$$J_f(v) := \sqrt{\sum_{m \in \mathbb{Z}^q} f(m) |\hat{v}(m)|^2}. \quad (\text{B.13})$$

We now characterize the extremal points of the unit ball of J_f in $L^2(T^q)$.

Lemma B.6. *Consider the set $B = \{v \in L^2_0(T^q) : J_f(v) \leq 1\}$. We have that*

$$\text{Ext}(B) = \{v \in L^2_0(T^q) : J_f(v) = 1\}. \quad (\text{B.14})$$

Proof. First note that $J_f(v) > 0$ whenever $v \neq 0$ thanks to Assumption *i*) of f . Setting

$$\mathcal{B} = \{v \in L^2_{\circ}(T^q) : J_f(v) = 1\}, \quad (\text{B.15})$$

we want to prove that $\text{Ext}(B) = \mathcal{B}$. We first proceed to prove the inclusion $\mathcal{B} \subset \text{Ext}(B)$. Given $v \in \mathcal{B}$ consider the convex combination

$$v = \lambda v_1 + (1 - \lambda)v_2 \quad (\text{B.16})$$

with $\lambda \in (0, 1)$ and $v_1, v_2 \in B$. Applying the functional J_f to both sides of (B.16), using the convexity of J_f and that $v_1, v_2 \in B$ we obtain that $\lambda J_f(v_1) + (1 - \lambda)J_f(v_2) = 1$ implying that $J_f(v_1) = J_f(v_2) = 1$. Therefore, from (B.16) we get

$$1 = \lambda^2 \sum_{m \in \mathbb{Z}^q} f(m) |\hat{v}_1(m)|^2 + (1 - \lambda)^2 \sum_{m \in \mathbb{Z}^q} f(m) |\hat{v}_2(m)|^2 + 2\lambda(1 - \lambda) \text{Re} \sum_{m \in \mathbb{Z}^q} f(m) \overline{\hat{v}_1(m)} \hat{v}_2(m)$$

implying, since $\lambda \neq 0$ and $\lambda \neq 1$, that

$$\text{Re} \sum_{m \in \mathbb{Z}^q} f(m) \overline{\hat{v}_1(m)} \hat{v}_2(m) = 1.$$

Recalling that $J_f(v_1) = J_f(v_2) = 1$, a simple contradiction argument that uses the sharpness of Cauchy-Schwarz inequality shows that $\hat{v}_1(m) = \hat{v}_2(m)$ for every $m \in \mathbb{Z}^q$. In particular, $v_1 = v_2 = v$ implying that v is an extremal point for B .

Conversely, let us show that $\text{Ext}(B) \subset \mathcal{B}$. Suppose that such inclusion does not hold true. Then, there exists $v \in \text{Ext}(B)$ such that $J_f(v) \neq 1$. In particular, since $v \in B$, there exists $\varepsilon > 0$ such that $J_f(v) < 1 - \varepsilon$. Consider $w \in L^2_{\circ}(T^q) \setminus \{0\}$ such that $J_f(w) \leq \frac{\varepsilon}{2}$. Notice that such w exists since $\text{dom } J_f \neq \{0\}$ and J_f is positively one-homogeneous. Consider then the convex combination

$$v = \frac{1}{2}(v + w) + \frac{1}{2}(v - w). \quad (\text{B.17})$$

Since $w \neq 0$, we immediately have that $v + w \neq v - w$. Moreover

$$J_f(v \pm w) \leq J_f(v) + J_f(w) < 1, \quad (\text{B.18})$$

showing that v is not an extremal point. \square

If f is strictly positive, then we can characterize the extremal points of J_f in $L^2(T^q)$ as follows.

Lemma B.7. *Suppose that f satisfies*

i') For every $m \in \mathbb{Z}^q$ it holds that $f(m) > 0$.

Then given the set $\tilde{B} = \{v \in L^2(T^q) : J_f(v) \leq 1\}$ we have that

$$\text{Ext}(\tilde{B}) = \{v \in L^2(T^q) : J_f(v) = 1\}. \quad (\text{B.19})$$

Since the proof follows closely the one of Lemma B.6 we omit it.