SpezialForschungsBereich F 32

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

Inverse problems in spaces of measures

Kristian Bredies Hanna K. Pikkarainen

SFB-Report No. 2010–037 November 2010

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the
Austrian Science Fund (FWF)

SFB sponsors:

- **Austrian Science Fund** (FWF)
- **University of Graz**
- **Graz University of Technology**
- **Medical University of Graz**
- **Government of Styria**
- **City of Graz**

Inverse problems in spaces of measures

Kristian Bredies* Hanna K. Pikkarainen†

November 15, 2010

Abstract

The ill-posed problem of solving linear equations in the space of vector-valued finite Radon measures with Hilbert-space data is considered. The well-posedness of regularization by minimizing the Tikhonov functional is established and further regularization properties are studied. In particular, the convergence rate of $\mathcal{O}(\delta)$ for the Bregman distance under a source condition is obtained. Furthermore, a flexible numerical minimization algorithm is proposed which converges subsequentially in the weak* sense and with $\mathcal{O}(n^{-1})$ in terms of the functional values. Finally, numerical results for sparse deconvolution demonstrate the applicability for a finite-dimensional discrete data space and infinite-dimensional solution space.

Keywords: Inverse problems, vector-valued finite Radon measures, Tikhonov regularization theory, delta-peak solutions, generalized conditional gradient method, iterative soft-thresholding, sparse deconvolution.

AMS Subject Classification (2010): 65J20, 46E27, 49M05.

1 Introduction

This paper is concerned with solving ill-posed linear inverse problems for measure-valued unknowns. While the regularization theory of inverse problems in Hilbert spaces is well-established [20], problems set in Banach spaces gained recent interest [4, 31, 36, 41, 30, 9, 42, 21, 46]. One particular setting, which has been in the focus of research, is regarding the recovery of sparse unknows, i.e., exploit the knowledge that the sought object is, in some sense, composed of only a few elements. This is usually formulated in the discrete setting imposing an ℓ^0 constraint [10, 16, 15, 25]. Relaxing this constraint

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

†hanna.pikkarainen@ricam.oeaw.ac.at, Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Altenbergerstraße 69, A-4040 Linz, Austria

leads to the so-called "sparse regularization" which is penalizing with an ℓ^1 functional. This approach gives inverse problems in sequence spaces which have been extensively studied in terms of regularization and numerical algorithms [14, 27, 35, 3, 12, 7, 8, 6, 28, 22, 19, 33]. In practical applications, the respective results and methods are transferred to the continuous setting without further justification, where they have been observed to work well [45, 17, 14, 43]. The present work aims at closing this theoretical gap by establishing a framework for this continuous setting in terms of studying the space of Radon measures as the solution space for linear inverse problems.

We consider linear inverse problems of the following type: Find μ such that

$$K^*\mu = f^0 \tag{1}$$

for some data $f^0 \in H$. Here, H is a Hilbert space which models the measured data and K^* maps continuously from the space of finite vector-valued Radon measures $\mathcal{M}(\Omega, \mathbb{R}^m)$ into H. The set Ω is allowed to be a continuum. As it will turn out, it is meaningful to consider linear mappings K^* which are the adjoint of some linear and continuous $K : H \to \mathcal{C}_0(\Omega, \mathbb{R}^m)$. In particular, the space of Radon measures allows for finite combinations of delta peaks, which are the continuous analogue of sparsity.

Solving the inverse problem (1) is often ill-posed in the sense that small variations in data cause large effects on the value of an unknown. For that reason, solution methods for inverse problems require stabilization. This can be done, for example, by using a regularization method where an ill-posed inverse problem is replaced by a well-posed problem. For our purposes, it is important that this stabilized well-posed problem respects the requirement for sparse solutions. This will be realized by penalizing with the norm in $\mathcal{M}(\Omega, \mathbb{R}^m)$ and solving the Tikhonov minimization problem

$$\min_{\mu \in \mathcal{M}(\Omega, \mathbb{R}^m)} \frac{\|K^*\mu - f^\delta\|^2}{2} + \alpha\|\mu\|_{\mathcal{M}} \tag{2}$$

where the linear and continuous $K : H \to \mathcal{C}_0(\Omega, \mathbb{R}^m)$ is the "predual" forward mapping, $f^\delta \in H$ the given noisy data corresponding to noise level $\delta > 0$ and $\alpha > 0$ a regularization parameter. This work studies existence of minimizers for this functional, optimality conditions and regularization properties. Moreover, a numerical algorithm is proposed which converges in the weak* sense.

A similar framework with Radon measures has been recently introduced in [43] which also includes some numerical experiments for the recovery of sparse unknowns. The authors reduce the situation to the Hilbert space setting by utilizing suitable embeddings of $\mathcal{M}(\Omega)$ into a Sobolev space of negative order, implying some dependence on the dimension. Our approach is genuinely set in the space of Radon measures, dimension-independent and covers the inverse problems considered in [43]. Moreover, differently from the discrete soft-thresholding algorithm used in the latter paper, we introduce and carry out the convergence analysis for an algorithm acting in $\mathcal{M}(\Omega, \mathbb{R}^m)$. In [11], elliptic partial differential equations are controlled with a Radon measure penalty. The authors establish a similar predual formulation and optimality system, we will obtain these results as a special case; namely with K^* in (2) being the solution operator of an elliptic equation. Finally, we mention that in [37], also Radon measures are the solution model for the inverse problem of deconvolution, regularization, however, is not performed in terms of Tikhonov functionals, but with an orthogonal matching pursuit.

Possible applications for the Radon-measure framework could be the deconvolution of point-like objects, a problem which arises, for instance, when long-exposure images of astronomical objects are taken with a ground-based telescope [39] or when mass spectrograms are acquired in which a substance contains multiple isotopes [5]. Moreover, it might be applied to formulate and solve the problem of finding an optimal placement for control devices or point actuators for partial differential equations [44].

The outline of the paper is as follows. Section 2 introduces the functional-analytic framework for solutions of (2). As a main result we show that this problem is the Fenchel dual for a minimization problem set in H involving the predual operator K mapping into $\mathcal{C}_0(\Omega, \mathbb{R}^m)$. Based on this, necessary and sufficient optimality conditions for minimizers of the functional are derived. In Section 3, convergence of the respective solutions of (2) as $\delta \to 0$ is studied. We derive the rate $\mathcal{O}(\delta)$ for the Bregman distance with respect to the norm in $\mathcal{M}(\Omega, \mathbb{R}^m)$ under the usual source condition. Due to the forward operator being an adjoint mapping, the source condition can be studied in $\mathcal{C}_0(\Omega, \mathbb{R}^m)$ instead of $\mathcal{M}(\Omega, \mathbb{R}^m)^*$. Section 4 proposes a practical numerical algorithm which produces a sequence of approximations each consisting of finitely many delta peaks. It is shown that this sequence minimizes the functional with rate $\mathcal{O}(n^{-1})$ and converges subsequentially in the weak* sense to a minimizer. The applicability of this algorithm is established in Section 5. There, a deconvolution problem is solved numerically in a semi-discrete setting with discrete data points but arbitrary peak positions. The article concludes with Section 6, summarizing the results.

2 Minimization in spaces of measures

This section is devoted to the analysis of the minimization problem (2). Before going into the precise definition of the setting, we introduce some notation and recall some facts from measure theory. We are dealing with vectors in $\mathbb{R}^m$ for $m \geq 1$ which we equip with a norm $|\cdot|$ and the associated dual norm $|\cdot|_*$ with respect to the dual pairing induced by the scalar product $\sigma \cdot \tau = \sum_{i=1}^m \sigma_i \tau_i$.

For the measure-theoretic part, we follow the notation of [2]. Let Ω

always be a separable locally compact metric space. A *finite Radon measure* μ is a σ-additive $\mathbb{R}^m$-valued set function defined on the Borel sets of Ω. Its *total-variation measure*, i.e.,

$$|\mu|(E) = \sup \Big\{ \sum_{n=1}^{\infty} |\mu(E_n)| \ : \ \bigcup_{n=1}^{\infty} E_n = E \ , \ \{E_n\} \text{ disjoint and measurable} \Big\}$$

is always a finite Radon measure. We denote the space of finite Radon measures by $\mathcal{M}(\Omega, \mathbb{R}^m)$ (and $\mathcal{M}(\Omega)$ if $m = 1$). It can be normed by

$$\|\mu\|_{\mathcal{M}} = |\mu|(\Omega) \ .$$

Each finite Radon measure μ possesses a *polar decomposition*, i.e., there exists a function $\sigma \in L^1_{|\mu|}(\Omega, \mathbb{R}^m)$ with $|\sigma(x)| = 1$ for $|\mu|$-almost all $x \in \Omega$ such that

$$\mu(E) = \int_E \sigma \ \mathrm{d}|\mu| \quad \text{for all measurable } E \subset \Omega \ .$$

An important result about this space, which is known as the Riesz theorem, is its characterization as a dual space. To state this theorem, introduce the space

$$\mathcal{C}_0(\Omega, \mathbb{R}^m) = \overline{\{u \in \mathcal{C}(\Omega, \mathbb{R}^m) \ : \ \operatorname{supp} u \subset\subset \Omega\}}_{\|\cdot\|_\infty} \quad , \quad \|u\|_\infty = \sup_{x \in \Omega} \ |u(x)|_*$$

which is the completion of continuous compactly supported functions with respect to the supremum norm. The Riesz theorem now states the following:

Theorem 1. [2, Theorem 1.54] *Let $L : \mathcal{C}_0(\Omega, \mathbb{R}^m) \to \mathbb{R}$ be linear and continuous. Then, there exists a unique finite Radon measure $\mu \in \mathcal{M}(\Omega, \mathbb{R}^m)$ such that L can be expressed by*

$$L(u) = \sum_{i=1}^{m} \int_\Omega u_i \ \mathrm{d}\mu_i \quad \text{for all } u \in \mathcal{C}_0(\Omega, \mathbb{R}^m) \ .$$

In particular, $\|L\| = \|\mu\|_{\mathcal{M}}$.

In other words, $\big(\mathcal{C}_0(\Omega, \mathbb{R}^m)\big)^* = \mathcal{M}(\Omega, \mathbb{R}^m)$ which also implies the completeness of the space of Radon measures. We therefore say that a sequence $\{\mu^n\}$ in $\mathcal{M}(\Omega, \mathbb{R}^m)$ converges in the weak* sense to a μ if and only if

$$\sum_{i=1}^{m} \int_\Omega u_i \ \mathrm{d}\mu_i^n \longrightarrow \sum_{i=1}^{m} \int_\Omega u_i \ \mathrm{d}\mu_i \quad \text{for each } u \in \mathcal{C}_0(\Omega, \mathbb{R}^m) \ .$$

Note that $\mu \mapsto \|\mu\|_{\mathcal{M}}$ is lower semi-continuous in this weak* sense. Moreover, the following compactness result holds true:

Theorem 2. [2, Theorem 1.59] *Let $\{\mu^n\}$ be a bounded sequence in $\mathcal{M}(\Omega, \mathbb{R}^m)$. Then, there exists a $\mu \in \mathcal{M}(\Omega, \mathbb{R}^m)$ such that μ is the weak* limit of a subsequence of $\{\mu^n\}$.*

Remark 1. Let $\Omega \subset \mathbb{R}^d$ be an open, non-empty set and denote by $|\cdot|_2$ the Euclidean norm in $\mathbb{R}^d$. It is clear that $\mathcal{C}_0(\Omega, \mathbb{R}^m) \hookrightarrow L^2(\Omega, \mathbb{R}^m)$ is dense, but unfortunately, this does not lead to the strong density of $L^2(\Omega, \mathbb{R}^m)$ in $\mathcal{M}(\Omega, \mathbb{R}^m)$ due to the lack of reflexivity. However, the density in the weak* sense holds, meaning that for each $\mu \in \mathcal{M}(\Omega, \mathbb{R}^m)$, there exists a sequence $\{w^n\}$ in $L^2(\Omega, \mathbb{R}^m)$ such that $w^n \mathcal{L}^d \rightharpoonup^* \mu$ with $\mathcal{L}^d$ being the d-dimensional Lebesgue measure.

This can be seen as follows. Construct the operators $M_\varepsilon : L^2(\Omega, \mathbb{R}^m) \to \mathcal{C}_0(\Omega, \mathbb{R}^m)$ by $M_\varepsilon : w \mapsto (\chi_{\Omega_\varepsilon} w) * G_{\varepsilon/2}$ where $\Omega_\varepsilon = \{x \in \Omega \ : \ \mathrm{dist}(x, \partial\Omega) > \varepsilon\}$ and $G_\varepsilon(x) = \varepsilon^{-d} G(x/\varepsilon)$ with $G \in \mathcal{C}_0^\infty(\Omega)$ being a standard mollifier. It is well-known that each M_ε is continuous.

Consider a $u \in \mathcal{C}(\Omega, \mathbb{R}^m)$ with compact support. The function u is uniformly continuous, thus, choose for $\delta > 0$ given an $\varepsilon > 0$ such that $|u(x) - u(y)|_* \leq \delta / \|G\|_1$ whenever $|x - y|_2 \leq \varepsilon$ where $|\cdot|_2$ is the Euclidean norm in $\mathbb{R}^m$. Moreover, one can choose ε small enough such that $\chi_{\Omega_\varepsilon} u = u$ and so $M_\varepsilon u = u * G_{\varepsilon/2}$. For each $x \in \Omega$ follows

$$\begin{aligned} |u(x) - M_\varepsilon u(x)|_* &= \Big| \int_{\{|x-y|_2 \leq \varepsilon/2\}} \big(u(y) - u(x)\big) G_{\varepsilon/2}(x - y) \ \mathrm{d}y \Big|_* \\ &\leq \Big(\sup_{\{|x-y|_2 \leq \varepsilon/2\}} |u(x) - u(y)|_* \Big) \int_{\mathbb{R}^d} |G_{\varepsilon/2}(x - y)| \ \mathrm{d}y \\ &\leq \delta. \end{aligned}$$

Hence $\|u - M_\varepsilon u\|_\infty \to 0$ as $\varepsilon \to 0$.

Now each $M_\varepsilon^* : \mathcal{M}(\Omega, \mathbb{R}^m) \to L^2(\Omega, \mathbb{R}^m)$ is continuous and yields, for $\mu \in \mathcal{M}(\Omega, \mathbb{R}^m)$, a sequence $w^\varepsilon = M_\varepsilon^* \mu$ such that, for each $u \in \mathcal{C}(\Omega, \mathbb{R}^m)$ with compact support,

$$\langle w^\varepsilon, u \rangle_2 = \sum_{i=1}^m \int_\Omega u(x) w_i^\varepsilon(x) \ \mathrm{d}x = \langle w^\varepsilon \mathcal{L}^d, u \rangle.$$

As well as

$$\langle w^\varepsilon \mathcal{L}^d, u \rangle = \langle \mu, M_\varepsilon u \rangle \to \langle \mu, u \rangle \quad \text{as } \varepsilon \to 0$$

implying the assertion $w^\varepsilon \mathcal{L}^d \rightharpoonup^* \mu$ since convergence was checked against a strongly dense subset of $\mathcal{C}_0(\Omega, \mathbb{R}^m)$.

Now, let us define more precisely the setting for the inverse problem (1). We consider the inversion of K^* which is the adjoint of a continuous mapping $K : H \to \mathcal{C}_0(\Omega, \mathbb{R}^m)$. Such an adjoint maps continuously between $\mathcal{M}(\Omega, \mathbb{R}^m)$ and H. It moreover is weak*-to-weak (sequentially) continuous which becomes important, as we will see, for showing weak* lower semi-continuity of associated Tikhonov functionals.

Remark 2. The requirement that the forward operator $A = K^*$ is an adjoint mapping is not a severe restriction if A is supposed to be weak*-to-weak(*) sequentially continuous: Recall [13, Thm. V.1.3.] that if X is a separable Banach space and a functional $u^{**} \in X^{**}$ is sequentially weak* continuous, then there has to be a $u \in X$ such that $\langle u^{**}, u^* \rangle = \langle u^*, u \rangle$ for all $u^* \in X^*$ (note that weak* continuity of a linear functional is equivalent with continuity with respect to the bounded X topology in X^* [18, Theorem V.5.6] which in turn can be characterized by sequential weak* continuity). Now for separable Banach spaces X, Y holds: an operator $A \in \mathcal{L}\big(Y^*, X^*\big)$ which is weak*-to-weak* sequentially continuous has an adjoint $A^* \in \mathcal{L}\big(X^{**}, Y^{**}\big)$ which maps X into Y regarded as closed subspaces in X^{**} and Y^{**} under the canonical injection into the bidual space, respectively. This follows from the fact that for $u \in X$ and each sequence $\{u^n\} \in Y^*$ with $v^n \rightharpoonup^* v$ for some $v \in Y^*$ we have

$$\lim_{n\to\infty} \langle A^*u,\, v^n \rangle = \lim_{n\to\infty} \langle u,\, Av^n \rangle = \langle u,\, Av \rangle = \langle A^*u,\, v \rangle$$

yielding that A^*u is weak* sequentially continuous on Y^* and is therefore representable by an element in Y. Hence, A^* can be considered as $A^* = K$ with $K \in \mathcal{L}\big(X, Y\big)$ and consequently, $A = K^*$ is an adjoint mapping.

Throughout this section, we denote by f some given data in H. It can be either the exact data f^0 or the noisy data f^δ. We use the notation T_α for the Tikhonov functional

$$T_\alpha(\mu) = \frac{\|K^*\mu - f\|^2}{2} + \alpha\|\mu\|_{\mathcal{M}}.$$

The following proposition shows that the Tikhonov functional has a minimizer.

Proposition 1. *Let $K : H \to \mathcal{C}_0(\Omega, \mathbb{R}^m)$ be a linear and continuous mapping, $f \in H$ and $\alpha > 0$. Then, the minimization problem*

$$\min_{\mu \in \mathcal{M}(\Omega, \mathbb{R}^m)} \frac{\|K^*\mu - f\|^2}{2} + \alpha\|\mu\|_{\mathcal{M}} \tag{3}$$

admits a solution $\mu^* \in \mathcal{M}(\Omega, \mathbb{R}^m)$. If K^* is injective, this solution is unique.

Proof. Observe that T_α is proper and coercive on $\mathcal{M}(\Omega, \mathbb{R}^m)$. It is moreover sequentially weak* lower semi-continuous: A sequence $\{\mu^n\}$ converging to μ in the weak* sense is mapped via K^* to a weakly convergent sequence $K^*\mu^n \rightharpoonup K^*\mu^n$ in H since

$$\langle K^*\mu^n,\, v \rangle = \sum_{i=1}^m \int_\Omega Kv \;\mathrm{d}\mu_i^n \to \sum_{i=1}^m \int_\Omega Kv \;\mathrm{d}\mu_i = \langle K^*\mu,\, v \rangle \quad \text{for all } v \in H\ .$$

Denote by $T_\alpha(\mu) = S(\mu) + \alpha R(\mu)$ with

$$S(\mu) = \tfrac{1}{2}\|K^*\mu - f\|^2 \qquad \text{and} \qquad R(\mu) = \|\mu\|_{\mathcal{M}}. \tag{4}$$

The mapping $w \mapsto \frac{1}{2}\|w - f\|^2$ is convex and continuous, and hence sequentially weak lower semi-continuous. This yields the sequential weak* lower semi-continuity of S. The functional R is just the norm in $\mathcal{M}(\Omega, \mathbb{R}^m)$ which is known to the sequentially lower semi-continuous in the weak* sense, implying the desired property for T_α.

By the direct method, a minimizing argument μ^* in $\mathcal{M}(\Omega, \mathbb{R}^m)$ exists. Finally, an injective K^* results in a strictly convex T_α which immediately gives the stated uniqueness. □

Before proceeding to the analysis of minimizers of (3), we introduce some examples which show what kind of problems can be treated within this framework. The first regards a deconvolution problem while the second involves elliptic partial differential equations.

Example 1. Let $\Omega \subset \mathbb{R}^d$ be a bounded domain and $k \in L^2(\mathbb{R}^d)$ with essential support in $\Omega' \subset \mathbb{R}^d$, a bounded open set. Denote by $\Omega'' = \{x - y \ : \ x \in \Omega\ ,\ y \in \Omega'\} = \Omega - \Omega'$ and define

$$(Kv)(x) = \int_\Omega k(y - x)v(y)\ \mathrm{d}y$$

for $x \in \Omega''$. We like to show that $Kv \in \mathcal{C}_0(\Omega'')$ by approximation with compactly supported functions. First, we see that the translation of a L^2-function, i.e., $k \mapsto k(\,\cdot - x)$, is a continuous operation mapping $\Omega'' \to L^2(\mathbb{R}^d)$ (see, for instance, [1, Theorem 2.32]) and restricting to Ω followed by taking the scalar product is also. Hence Kv is indeed a continuous function. Now set $\Omega_\varepsilon = \{x \in \Omega \ : \ \mathrm{dist}(x, \partial\Omega) > \varepsilon\}$ and let $v^\varepsilon = \chi_{\Omega_\varepsilon} v$. Clearly, $v^\varepsilon \to v$ in $L^2(\Omega)$ as $\varepsilon \to 0$. Construct k^ε analogously and define $K_\varepsilon v^\varepsilon$ according to the above with v and k replaced by v^ε and k^ε, respectively.

For each $\varepsilon > 0$, we deduce the inclusion $\operatorname{supp} K_\varepsilon v^\varepsilon \subset \overline{\Omega_\varepsilon - \Omega'_\varepsilon} \subset\subset \Omega''$. Therefore, let $\{z^n\}$ with $z^n \in \Omega_\varepsilon - \Omega'_\varepsilon$ be an arbitrary sequence, so $z^n = x^n - y^n$ with each $x^n \in \overline{\Omega_\varepsilon} \subset\subset \Omega$ and $y^n \in \overline{\Omega'_\varepsilon} \subset\subset \Omega'$. Therefore, there exist a respective subsequences, not relabeled, and limits $x \in \overline{\Omega_\varepsilon}$, $y \in \overline{\Omega'_\varepsilon}$ such that $x^n \to x$ and $y^n \to y$ as $n \to \infty$. Hence $z = x - y \in \overline{\Omega_\varepsilon} - \overline{\Omega'_\varepsilon} = \overline{\Omega_\varepsilon - \Omega'_\varepsilon} \subset \Omega''$ (note that both sets are compact) is a limit of $\{z^n\}$ in Ω'', which means that the claimed compactness holds.

Thus, $K_\varepsilon v^\varepsilon$ is continuous with compact support in Ω''. The bilinear operation

$$L(v,k)(x) = \int_\Omega k(y - x)v(y)\ \mathrm{d}y$$

obeys, because of the Cauchy-Schwarz inequality,

$$\|L(v,k)\|_\infty \le \|k\|_2\|v\|_2$$

which means that $L : L^2(\Omega) \times L^2(\Omega') \to \mathcal{C}_0(\Omega'')$ is a continuous mapping. Therefore, we have $L(v^\varepsilon, k^\varepsilon) \to Kv$ as $\varepsilon \to 0$ which implies $Kv \in \mathcal{C}_0(\Omega'')$ with K being continuous.

Hence, Proposition 1 is applicable with $H = L^2(\Omega)$ and K mapping into $\mathcal{C}_0(\Omega'')$. Let us eventually compute $K^* : \mathcal{M}(\Omega'') \to L^2(\Omega)$: For $\mu \in \mathcal{M}(\Omega'')$

$$\langle K^*\mu,\, v\rangle = \int_{\Omega''} \int_{\Omega} k(y-x)v(y)\ \mathrm{d}y\ \mathrm{d}\mu(x) = \int_{\Omega} \int_{\Omega''} k(y-x)\ \mathrm{d}\mu(x)v(y)\ \mathrm{d}y$$

by Fubini's theorem, so $K^*\mu = \mu * k$ in the sense of the convolution of measures. It turns out that (3) is actually the deconvolution problem

$$\min_{\mu \in \mathcal{M}(\Omega - \Omega')} \frac{\|\mu * k - f\|_2^2}{2} + \alpha \|\mu\|_{\mathcal{M}}$$

where the sought solution is a measure in $\Omega - \Omega'$ and the data is compared only in Ω.

Example 2. Let $\Omega \subset \mathbb{R}^d$, $d \leq 3$, be a bounded domain with a $\mathcal{C}^{1,1}$-boundary and $A \in \mathcal{C}^{0,1}(\overline{\Omega}, \mathbb{R}^{d\times d})$ a matrix field such that $A(x)$ is symmetric for each $x \in \Omega$ and there is a uniform ellipticity constant $c > 0$ such that $\xi^{\mathrm{T}} A(x)\xi \geq c\sum_{i=1}^{d} \xi_i^2$ for each $x \in \Omega$ and $\xi \in \mathbb{R}^d$. Consider K as the weak solution operator $K : v \mapsto u$ such that

$$\begin{aligned} -\operatorname{div}(A\nabla u) &= v \quad \text{in } \Omega\ , \\ u &= 0 \quad \text{on } \partial\Omega\ . \end{aligned} \tag{5}$$

From the regularity theory of elliptic equations (see, for instance, [29, Chapter 2]) we know that $K : L^2(\Omega) \to H_0^1(\Omega) \cap H^2(\Omega)$ continuously. Since $d \leq 3$, the solution has to be continuous due to the continuous embedding $H^2(\Omega) \hookrightarrow \mathcal{C}(\overline{\Omega})$. Furthermore, $u \in H_0^1(\Omega)$ implies a zero trace in $H^{1/2}(\partial\Omega)$ which in particular gives a zero trace for continuous functions in $H_0^1(\Omega)$. Hence, $u \in \mathcal{C}_0(\Omega)$ holds and consequently $K : L^2(\Omega) \to \mathcal{C}_0(\Omega)$ is continuous, implying that Proposition 1 is applicable.

The adjoint $K^* : \mathcal{M}(\Omega) \to L^2(\Omega)$ is the solution operator $K^* : \mu \mapsto f$ for the elliptic boundary value problem

$$\begin{aligned} -\operatorname{div}(A\nabla f) &= \mu \quad \text{in } \Omega\ , \\ f &= 0 \quad \text{on } \partial\Omega \end{aligned} \tag{6}$$

where the boundary condition needs to be interpreted properly.

This can be seen as follows. Let $\mu \in \mathcal{M}(\Omega)$. If μ_ε converges in weak* sense to μ, by the properties of an adjoint $K^*\mu_\varepsilon$ converges weakly in $L^2(\Omega)$ to $K^*\mu$. By Remark 1, there exists a sequence $\{w_\varepsilon\}$ in $L^2(\Omega)$ such that $w_\varepsilon \mathcal{L}^d \rightharpoonup^* \mu$ as $\varepsilon \to 0$. Therefore $K^*\mu$ can be defined by

$$K^*\mu = \operatorname*{w-lim}_{\varepsilon\to 0}\ K^*(w_\varepsilon \mathcal{L}^d)$$

since the weak limit exists and does not depend on w_ε. By the definition,

$$\langle K^*(w_\varepsilon \mathcal{L}^d),\, v\rangle_2 = \langle w_\varepsilon \mathcal{L}^d,\, Kv\rangle = \langle w_\varepsilon,\, Kv\rangle_2 \quad \text{for all } v \in L^2(\Omega)\ .$$

Let $v \in L^2(\Omega)$ be arbitrary. Using Green's formula [29, Lemma 1.5.3.2], the boundary conditions and the symmetry of A,

$$\begin{aligned}\langle Kw_\varepsilon,\, v\rangle_2 &= -\int_\Omega Kw_\varepsilon \operatorname{div}(A\nabla Kv)\ \mathrm{d}x \\ &= -\int_\Omega \operatorname{div}(A\nabla Kw_\varepsilon)Kv\ \mathrm{d}x = \langle w_\varepsilon,\, Kv\rangle_2\ .\end{aligned}$$

Thus $K^*(w_\varepsilon \mathcal{L}^d) = Kw_\varepsilon$, i.e., $K^*(w_\varepsilon \mathcal{L}^d)$ is the weak solution of (5) with the source function w_ε. Hence the solution of the elliptic problem (6) can be defined to be the weak L^2-limit of the solutions of (5) with the source functions w_ε where $w_\varepsilon \mathcal{L}^d \rightharpoonup^* \mu$ as $\varepsilon \to 0$. Since each Kw_ε has a zero trace, the boundary condition in (6) is justified even though the trace for L^2-functions is not defined in the usual sense.

Next, we will derive optimality conditions for minimizers of (3). The usual convex analysis, however, which requires the consideration of the dual space $\mathcal{M}(\Omega, \mathbb{R}^m)^*$, turns out to be complicated due to the inaccessibility of this space. There is, nevertheless, a method which circumvents these complications: One can identify (3) being equivalent to the Fenchel dual problem associated with a functional in H. The corresponding optimality conditions only involve the dual space $\mathcal{C}_0(\Omega, \mathbb{R}^m)^*$, the space of finite Radon measures.

Proposition 2. *Let $K : H \to \mathcal{C}_0(\Omega, \mathbb{R}^m)$ be linear and continuous, $f \in H$ and $\alpha > 0$. The problem*

$$\min_{v\in H} \frac{\|v - f\|^2}{2} + I_{\{v\in H\ :\ \|Kv\|_\infty \le \alpha\}}(v) \tag{7}$$

is equivalent to a Fenchel predual problem of (3)*, i.e., the Fenchel dual of* (7) *exists and possesses the same minimizers as* (3)*. Here, the indicator functional I of a set $E \subset \mathcal{C}_0(\Omega, \mathbb{R}^m)$ is defined by $I_E(v) = 0$ if $v \in E$ and $I_E(v) = \infty$ if $v \notin E$.*

Proof. Define $S_0(v) = \frac{1}{2}\|v - f\|^2$ for $v \in H$ whose conjugate function obeys

$$S_0^*(w) = \frac{\|f + w\|^2}{2} - \frac{\|f\|^2}{2}\ .$$

Likewise, the conjugate of $R_0 : \mathcal{C}_0(\Omega, \mathbb{R}^m) \to \mathbb{R} \cup \{\infty\}$ given by $R_0(u) = I_{\{\|u\|_\infty \le \alpha\}}(u)$ reads as

$$R_0^*(\mu) = \sup_{\|u\|_\infty \le \alpha} \sum_{i=1}^m \int_\Omega u_i\ \mathrm{d}\mu_i = \alpha\|\mu\|_{\mathcal{C}_0^*} = \alpha\|\mu\|_{\mathcal{M}} = \alpha R(\mu)$$

where R is defined in (4). Now verify that all prerequisites for the validity of the Fenchel duality formula [48, Corollary 2.8.5] are satisfied. In particular, $0 \in H$ is in the domain of $S_0 + R_0 \circ K$ and R_0 is continuous at $0 \in \mathcal{C}_0(\Omega, \mathbb{R}^m)$. Hence,

$$\min_{v \in H} \frac{\|v-f\|^2}{2} + I_{\{\|Kv\|_\infty \leq \alpha\}}(v) = \max_{\mu \in \mathcal{M}(\Omega, \mathbb{R}^m)} \frac{\|f\|^2}{2} - \frac{\|f - K^*\mu\|^2}{2} - \alpha \|\mu\|_{\mathcal{M}}$$

according to the Fenchel duality formula and the above considerations. The solutions for the problem on the right-hand side obviously coincide with the solutions of (3). □

With the help of the predual problem and duality, it is possible to derive optimality conditions for the original problem.

Proposition 3. *Let $K : H \to \mathcal{C}_0(\Omega, \mathbb{R}^m)$ be linear and continuous, $f \in H$ and $\alpha > 0$. Then, a $\mu^* \in \mathcal{M}(\Omega, \mathbb{R}^m)$ is a solution of (3) if and only if $w^* = -K(K^*\mu^* - f)$ satisfies $\|w^*\|_\infty \leq \alpha$ and each polar decomposition $\mu^* = \sigma^*|\mu^*|$ obeys*

$$\sigma^* \cdot w^* = \alpha \quad |\mu^*|\text{-almost everywhere} . \tag{8}$$

In particular, $\operatorname{supp} |\mu^*| \subset \{x \in \Omega \ : \ |w^*(x)|_* = \alpha\}$ *for each solution μ^*.*

Proof. From Fenchel duality and Proposition 2 one knows that $(v^*, \mu^*) \in H \times \mathcal{M}(\Omega, \mathbb{R}^m)$ are solutions of (7) and (3), respectively, if and only if $-K^*\mu^* \in \partial S_0(v^*)$ and $\mu^* \in \partial R_0(Kv^*)$ with S_0 and R_0 from the proof of Proposition 2. This means that $-K^*\mu^* = v^* - f$ and $\|Kv^*\|_\infty \leq \alpha$ with

$$\sum_{i=1}^m \int_\Omega (w - Kv^*)_i \, \mathrm{d}\mu_i^* \leq 0 \quad \text{for all } w \in \mathcal{C}_0(\Omega, \mathbb{R}^m) \ , \ \|w\|_\infty \leq \alpha \ .$$

Plugging $v^* = f - K^*\mu^*$ and using a polar decomposition of μ^*, one sees that μ^* being optimal is equivalent to $\|w^*\|_\infty \leq \alpha$ and

$$\int_\Omega (w - w^*) \cdot \sigma^* \, \mathrm{d}|\mu^*| \leq 0 \quad \text{for all } w \in \mathcal{C}_0(\Omega, \mathbb{R}^m) \ , \ \|w\|_\infty \leq \alpha \ . \tag{9}$$

Rearranging, using basic properties of the supremum and the definition of the dual norm yields that this is equivalent to $\|w^*\|_\infty \leq \alpha$ and

$$\alpha|\mu^*|(\Omega) = \alpha\|\mu^*\|_{\mathcal{M}} = \sup_{\|w\|_\infty \leq \alpha} \int_\Omega w \cdot \sigma^* \, \mathrm{d}|\mu^*| \leq \int_\Omega w^* \cdot \sigma^* \, \mathrm{d}|\mu^*| \ .$$

This is, in turn, by basic properties of the integral, equivalent to $\|w^*\|_\infty \leq \alpha$ and $w^* \cdot \sigma^* = \alpha$ $|\mu^*|$-almost everywhere, proving the desired characterization.

We finally show $\operatorname{supp} |\mu^*| \subset \{x \in \Omega \ : \ |w^*(x)|_* = \alpha\}$ for a solution μ^* indirectly. For this purpose, let $x \in \Omega$ be such that $|w^*(x)|_* < \alpha$. Since w^*

is continuous, this has to be true on a whole neighborhood V_x of x. For each $u \in \mathcal{C}_0(V_x, \mathbb{R}^m)$, a $t > 0$ can be chosen such that both $w = w^* \pm tu$ obey $\|w\|_\infty \leq \alpha$. Plugged into (9), this gives

$$\int_\Omega u \cdot \sigma^* \, \mathrm{d}|\mu^*| = 0 \quad \text{for all } u \in \mathcal{C}_0(V_x, \mathbb{R}^m)$$

implying that μ^* restricted to V_x is the zero-measure and consequently $V_x \subset \Omega \backslash \operatorname{supp} |\mu^*|$. □

Remark 3. If the vector norm $|\cdot|$ is strictly convex, i.e., $|u|, |v| \leq 1$ implies $|u + v| < 2$, then there is a polar decomposition of an optimal μ^* for which $\sigma^* \in \mathcal{C}_0(\Omega, \mathbb{R}^m)$. This can be seen as follows. First, observe that necessarily, $\sigma^*(x) = J\big(w^*(x)\big)$ for $|\mu^*|$-almost all $x \in \Omega$ where $J : \mathbb{R}^m \to \mathbb{R}^m$ is the $|\cdot|_*$-$|\cdot|$-duality mapping with gauge $\phi : t \mapsto t/\alpha$: Since $\operatorname{supp} |\mu^*| \subset \{|w^*(x)|_* = \alpha\}$, we only need to consider $|\mu^*|$-almost every $x \in \Omega$ with $|w^*(x)|_* = \alpha$ for which we have, due to (8),

$$\left.\begin{aligned} w^*(x) \cdot \sigma^*(x) = \alpha = |w^*(x)|_* \phi\big(|w^*(x)|_*\big) \\ |\sigma^*(x)| = 1 = \phi\big(|w^*(x)|_*\big) \end{aligned}\right\} \quad \Longrightarrow \quad \sigma^*(x) = J\big(w^*(x)\big)$$

noting that strict convexity in the dual space leads to a single-valued duality relation.

On the other hand, each Borel measureable σ^* coinciding with $J \circ w^*$ $|\mu^*|$-almost everywhere gives a polar decomposition since it always defines the same integral. Put in a different way: $\sigma^*(x) = J\big(w^*(x)\big)$ for $|\mu^*|$ almost every $x \in \Omega$ constitutes the angular part of a polar decomposition of μ^*. Finally, this construction yields an element in $\mathcal{C}_0(\Omega, \mathbb{R}^m)$ by the uniform continuity of J (note that strict convexity on finite dimensional spaces implies uniform convexity) as well as $J(0) = 0$.

Remark 4. In the special case of $\Omega = \mathbb{N}$ with the discrete topology and $m = 1$, the space $\mathcal{C}_0(\Omega)$ is the space c_0 of null sequences and $\mathcal{M}(\Omega) = \ell^1$. The associated regularized problem (3) amounts to the regularization of a linear inverse problem with the ℓ^1-norm which is known to implement so-called *sparsity constraints*. As already pointed out in the introduction, these types of problems have been studied extensively in the literature, usually with a linear and continuous operator $K^* : \ell^2 \to \ell^2$.

Optimality in this case is often expressed as: u^* interpreted as a sequence is optimal if and only if

$$u_k^* \begin{cases} = 0 & \Rightarrow \quad |K(K^*u^* - f)|_k \leq \alpha \\ > 0 & \Rightarrow \quad \big(-K(K^*u^* - f)\big)_k = \alpha \\ < 0 & \Rightarrow \quad \big(-K(K^*u^* - f)\big)_k = -\alpha \; . \end{cases}$$

The translation into each polar decomposition of the corresponding measure is just $\mu^* = \sigma^*|\mu^*|$ with $|\mu^*| = |u^*|\mathcal{H}^0$ where $\mathcal{H}^0$ denotes the counting measure on $\mathbb{N}$ and $\sigma_k^* \in \operatorname{sgn}(u_k^*)$, i.e., $|\sigma_k^*| \leq 1$ if $u_k^* = 0$ and $\sigma_k^* = u_k^*/|u_k^*|$ else. The support of $|\mu^*|$ is exactly the collection of coefficients where $u_k^* \neq 0$, or, the other way around, the set of coefficients where $u_k^* = 0$ is $|\mu^*|$-negligible. Hence, it is easy to see that condition (8) is an equivalent description of the above in terms of measures.

Remark 5. The last remark shows that the situation (3) can be interpreted as a natural generalization of the discrete setting $\Omega = \mathbb{N}$ to continuous ones. In fact, the consideration of measure spaces is necessary since they guarantee the topological properties (sequential weak* compactness) which are required for the application of the direct method.

In case of $\Omega \subset \mathbb{R}^d$ being a domain, one could be tempted with regularizing with the L^1-norm instead of the norm in $\mathcal{M}(\Omega, \mathbb{R}^m)$ (in which, nevertheless, $L^1(\Omega, \mathbb{R}^m)$ is a closed subspace). However, in this case, we are only able to obtain a bounded sequence in $L^1(\Omega, \mathbb{R}^m)$ which is not sufficient for the extraction of a weakly convergent subsequence (see the Dunford-Pettis theorem, for instance [2, Thm. 1.38]). Based on this, one can easily construct a counterexample for the existence of solutions for the Tichonov functional with L^1-regularization (with some technical effort). As we have seen in Proposition 1, considering $\mathcal{M}(\Omega, \mathbb{R}^m)$ instead of $L^1(\Omega, \mathbb{R}^m)$ provides the compactness needed in the existence proof.

3 Regularization properties

In this section we interpret the minimization problem (2) as a Tikhonov regularization method for the inverse problem (1). We show that the minimization problem (2) is well-posed and approximates the inverse problem (1) in the sense usually considered in regularization theory.

In the setup of the paper, i.e., studying the inverse problem (1) where $K : H \to \mathcal{C}_0(\Omega, \mathbb{R}^m)$, the crucial observations are

- the operator $K^* : \mathcal{M}(\Omega, \mathbb{R}^m) \to H$ is weak*-to-weak continuous,
- the norm $\|\cdot\|$ is lower semi-continuous with respect to the weak topology in H,
- the norm $\|\cdot\|_{\mathcal{M}}$ is weak* lower semi-continuous [2, Theorem 1.59],
- every bounded sequence in $\mathcal{M}(\Omega, \mathbb{R}^m)$ has a weak* convergent subsequence [2, Theorem 1.59].

In what follows, we denote by f^0 the exact data and by f^δ noisy data. We use also the notation $T_\alpha^\delta(\mu) := \frac{1}{2}\|K^*\mu - f^\delta\|^2 + \alpha\|\mu\|_{\mathcal{M}}$ and call T_α^δ a Tikhonov functional. By Proposition 1, the Tikhonov functional T_α^δ has a minimizer and the minimizer is unique if K^* is injective. Hence the well-posedness of the minimization of the Tikhonov functional T_α^δ is guaranteed by the following stability result.

Proposition 4. *Let $K : H \to \mathcal{C}_0(\Omega, \mathbb{R}^m)$ be a linear and continuous mapping, $f^\delta \in H$ and $\alpha > 0$. Let $\{f^n\}$ be a sequence converging to f^δ in $(H, \|\cdot\|)$. Then every sequence $\{\mu^n\}$ satisfying*

$$\mu^n \in \operatorname*{arg\,min}_{\mu \in \mathcal{M}(\Omega, \mathbb{R}^m)} \frac{\|K^*\mu - f^n\|^2}{2} + \alpha\|\mu\|_{\mathcal{M}}$$

has a subsequence which converges in the weak sense, and the limit of each weak* convergent subsequence is a minimizer of T_α^δ.*

Proof. The claim is a special case of [31, Theorem 3.2]. □

If the null space of the operator K^* is not trivial, the inverse problem (1) may have multiple solutions. Our interest is in the following special solutions:

Definition 1. An element $\mu^\dagger \in \mathcal{M}(\Omega, \mathbb{R}^m)$ is called a *minimum norm solution* of (1) if

$$\|\mu^\dagger\|_{\mathcal{M}} = \min\{\|\mu\|_{\mathcal{M}} : K^*\mu = f^0\} < \infty.$$

We call a minimizer of T_α^δ a Tikhonov regularized solution of (1). In the regularization theory, the amount of the noise in the data is often given in the form of the noise level δ, i.e., $\|f^\delta - f^0\| \leq \delta$. A Tikhonov regularized solution is a good approximation of a solution of the inverse problem (1) if there exists a choice of the regularization parameter α such that a Tikhonov regularized solution converges to a minimum norm solution as the noise level tends to zero. The following proposition contains a convergence result with an a-priori parameter choice rule.

Proposition 5. *Let $K : H \to \mathcal{C}_0(\Omega, \mathbb{R}^m)$ be a linear and continuous mapping, $f^0 \in H$ and $\alpha > 0$. Assume that there exists a solution of (1). Let $f^\delta \in H$ and $\|f^\delta - f^0\| \leq \delta$. Let $\alpha(\delta)$ be such that*

$$\alpha(\delta) \longrightarrow 0 \quad \textit{and} \quad \frac{\delta^2}{\alpha(\delta)} \longrightarrow 0 \quad \textit{as } \delta \longrightarrow 0$$

and $\alpha(\,\cdot\,)$ is monotonically increasing. Suppose that the sequence $\{\delta_n\}$ converges monotonically to zero. Then a sequence $\{\mu^n\}$ satisfying

$$\mu^n \in \operatorname*{arg\,min}_{\mu \in \mathcal{M}(\Omega, \mathbb{R}^m)} \frac{\|K^*\mu - f^{\delta_n}\|^2}{2} + \alpha(\delta_n)\|\mu\|_{\mathcal{M}}$$

has a weak convergent subsequence as $\delta_n \to 0$. The limit of each weak* convergent subsequence is a minimum norm solution of (1). If, in addition, the minimum norm solution $\mu^\dagger$ is unique, the whole sequence $\{\mu^n\}$ converges to $\mu^\dagger$ in the weak* topology as $\delta_n \to 0$.*

Proof. A minimum norm solution of (1) exists by [31, Theorem 3.4]. Furthermore, the claim is a corollary of [31, Theorem 3.5]. □

For obtaining qualitative estimates for the convergence of a Tikhonov regularized solution to a minimum norm solution, some additional assumptions are needed. Usually these assumptions are given in the form of a so-called source condition. In here, we follow the lines of [9]. A minimum norm solution $\mu^\dagger$ is said to fulfill a source condition if

$$\exists\, h \in H : \quad K^{**}h \in \partial\|\cdot\|_{\mathcal{M}}(\mu^\dagger). \tag{10}$$

The subdifferential of the norm $\|\cdot\|_{\mathcal{M}}$ is $\partial\|\cdot\|_{\mathcal{M}}(\mu) = \{\varphi \in \mathcal{M}^* : \langle\varphi,\mu\rangle = \|\mu\|_{\mathcal{M}}$ and $\|\varphi\|_{\mathcal{M}^*} = 1\}$. By the Hahn-Banach theorem, the subdifferential $\partial\|\cdot\|_{\mathcal{M}}(\mu)$ is nonempty for all $\mu \in \mathcal{M}(\Omega,\mathbb{R}^m)$. Hence the source condition is

$$\exists\, h \in H : \quad \langle K^{**}h, \mu^\dagger\rangle = \|\mu^\dagger\|_{\mathcal{M}} \quad \text{and} \quad \|K^{**}h\|_{\mathcal{M}^*} = 1.$$

Using the definition of the adjoint and the Hahn-Banach theorem, the source condition can be written equivalently in the form

$$\exists\, h \in H : \quad \langle\mu^\dagger, Kh\rangle = \|\mu^\dagger\|_{\mathcal{M}} \quad \text{and} \quad \|Kh\|_\infty = 1. \tag{11}$$

Remark 6. The source condition of the form (10) is related to the subdifferential of the penalty term and hence usually requires information about the dual space of the space of the unknown. For finite Radon measures the dual space $\mathcal{M}^*$ is complicated and therefore at first glance the source condition is hard to verify. In this paper, the form of the inverse problem (1), i.e., the existence of the predual operator K, simplifies the source condition, as has been seen in (11). For this particular problem the dual space $\mathcal{M}^*$ or its norm is not needed in the source condition but it is not the case for all inverse problems involving Radon measures.

By the polar decomposition of $\mu^\dagger$ and the definition of the total variation norm,

$$\langle\mu^\dagger, Kh\rangle = \|\mu^\dagger\|_{\mathcal{M}} \quad \Longleftrightarrow \quad \int_\Omega (1 - (Kh)(x)\cdot\sigma^\dagger(x))\,\mathrm{d}|\mu^\dagger|(x) = 0.$$

According to the source condition (11), $\|Kh\|_\infty = 1$. Furthermore, $|\sigma^\dagger(x)| = 1$ for $|\mu^\dagger|$-almost all $x \in \Omega$. Therefore the source element h has to fulfill $(Kh)(x)\cdot\sigma^\dagger(x) = 1$ for $|\mu^\dagger|$-almost all $x \in \Omega$. Hence the source condition can equivalently be given as

$$\exists\, h \in H : \quad \begin{cases} (Kh)(x)\cdot\sigma^\dagger(x) = 1, & \text{for } |\mu^\dagger|\text{-almost every } x \in \Omega, \\ |(Kh)(x)| \le 1, & x \in \Omega. \end{cases} \tag{12}$$

In the Banach space theory of variational regularization, convergence rates results are given using generalized Bregman distances.

Definition 2. Let U be a Banach space and $J : U \to [0, \infty]$ a convex, proper functional with the subdifferential ∂J. The Bregman distance related to J is defined as the set-valued function

$$D_J(u, v) = \{J(u) - J(v) - \langle \xi, u - v\rangle : \xi \in \partial J(v)\}, \quad u, v \in U.$$

For the norm $\|\cdot\|_{\mathcal{M}}$ the corresponding Bregman distance $D_{\|\cdot\|_{\mathcal{M}}}(\mu, \nu)$ is the nonempty set

$$D_{\|\cdot\|_{\mathcal{M}}}(\mu, \nu) = \{\|\mu\|_{\mathcal{M}} - \langle \varphi, \mu\rangle : \langle \varphi, \nu\rangle = \|\nu\|_{\mathcal{M}} \text{ and } \|\varphi\|_{\mathcal{M}^*} = 1\}$$

for all $\mu, \nu \in \mathcal{M}(\Omega, \mathbb{R}^m)$. In the following proposition a convergence rates result with an a-priori parameter choice rule is stated.

Proposition 6. *Let $K : H \to \mathcal{C}_0(\Omega, \mathbb{R}^m)$ be a linear and continuous mapping, $f^0 \in H$ and $\alpha > 0$. Assume that there exists a solution of* (1)*. Let $f^\delta \in H$ and $\|f^\delta - f^0\| \le \delta$. In addition, assume that a minimum norm solution $\mu^\dagger$ satisfies the source condition* (11)*. Then, for each minimizer μ_α^δ*

$$\mu_\alpha^\delta \in \operatorname*{arg\,min}_{\mu \in \mathcal{M}(\Omega, \mathbb{R}^m)} \frac{\|K^*\mu - f^\delta\|^2}{2} + \alpha\|\mu\|_{\mathcal{M}}$$

there exists $d \in D_{\|\cdot\|_{\mathcal{M}}}(\mu_\alpha^\delta, \mu^\dagger)$ such that the estimate

$$d \le \frac{1}{2}\alpha\|h\|^2 + \frac{\delta^2}{2\alpha}$$

holds. In particular, if $\alpha \sim \delta$,

$$d = \mathcal{O}(\delta).$$

Proof. A minimum norm solution of (1) exists by [31, Theorem 3.4]. The claim itself follows from [9, Theorem 2]. □

Remark 7. Let us now consider an inverse problem $K^*\mu = f$ where $\mu \in \mathcal{M}(\Omega)$ and $f \in H$. Let $\mu^\dagger$ be a minimum norm solution of the problem. By the polar decomposition there exists $\sigma^\dagger \in L^1_{|\mu^\dagger|}(\Omega)$ such that $|\sigma^\dagger(x)| = 1$ for $|\mu^\dagger|$-almost all $x \in \Omega$ and $\mathrm{d}\mu^\dagger = \sigma^\dagger \, \mathrm{d}|\mu^\dagger|$. Let us redefine the value of $\sigma^\dagger$ to be 1 in the $|\mu^\dagger|$-null set $\{x \in \Omega : |\sigma^\dagger(x)| \ne 1\}$. Then the space Ω can be divided into two disjoint parts; $\Omega_+^\dagger = \{x \in \Omega : \sigma^\dagger(x) = 1\}$ and $\Omega_-^\dagger = \{x \in \Omega : \sigma^\dagger(x) = -1\}$.

The minimum norm solution $\mu^\dagger$ fulfills the source condition (12) if and only if there exist sets $\widetilde{\Omega}_+ \subset \Omega_+^\dagger$ and $\widetilde{\Omega}_- \subset \Omega_-^\dagger$, and an element $h \in H$ such that

$$\begin{cases} |\mu^\dagger|(\Omega_+^\dagger \setminus \widetilde{\Omega}_+) = 0, \\ |\mu^\dagger|(\Omega_-^\dagger \setminus \widetilde{\Omega}_-) = 0, \end{cases} \quad \text{and} \quad \begin{cases} (Kh)(x) = 1, & x \in \widetilde{\Omega}_+, \\ (Kh)(x) = -1, & x \in \widetilde{\Omega}_-, \\ |(Kh)(x)| \le 1, & x \in \Omega. \end{cases} \tag{13}$$

Since Kh is a $\mathcal{C}_0(\Omega)$-function, in view of (13) the source condition requires that the sets $\widetilde{\Omega}_+$ and $\widetilde{\Omega}_-$ have to be separated from each others and the boundary $\partial\Omega$ everywhere by a positive distance in Ω.

Example 2. (continued) Consider now the elliptic boundary value problem (6). Let the assumptions made in Example 2 be valid. Let $f \in L^2(\Omega)$ and $\mu^\dagger \in \mathcal{M}(\Omega)$ be a minimum norm solution of the inverse problem $K^*\mu = f$ where K^* is the solution operator for the elliptic problem (6).

With the notations of Remark 7, let us suppose that there exist sets $\widetilde{\Omega}_+ \subset \Omega_+^\dagger$ and $\widetilde{\Omega}_- \subset \Omega_-^\dagger$ such that

$$\begin{cases} |\mu^\dagger|(\Omega_+^\dagger \setminus \widetilde{\Omega}_+) = 0, \\ |\mu^\dagger|(\Omega_-^\dagger \setminus \widetilde{\Omega}_-) = 0, \end{cases} \quad \text{and} \quad \begin{cases} \operatorname{dist}(\widetilde{\Omega}_+, \widetilde{\Omega}_-) > 0, \\ \operatorname{dist}(\widetilde{\Omega}_+, \partial\Omega) > 0, \\ \operatorname{dist}(\widetilde{\Omega}_-, \partial\Omega) > 0 \end{cases} \tag{14}$$

where $\operatorname{dist}(E, F)$ is the distance between sets E and F in $\mathbb{R}^d$ defined by

$$\operatorname{dist}(E, F) = \inf\{|x - y|_2 : x \in E,\ y \in F\}.$$

According to Remark 7, the existence of sets $\widetilde{\Omega}_+$ and $\widetilde{\Omega}_-$ satisfing requirements (14) is a necessary condition for $\mu^\dagger$ to fulfill the source condition (13).

Using, for instance, [32, Theorem 1.4.1], we can construct functions $g_+, g_- \in \mathcal{C}^\infty(\Omega)$ with disjoint compact supports such that $0 \le g_+, g_- \le 1$,

$$g_+(x) = \begin{cases} 1, & x \in \widetilde{\Omega}_+, \\ 0, & x \in \overline{\widetilde{\Omega}}_- \cup \partial\Omega, \end{cases} \quad \text{and} \quad g_-(x) = \begin{cases} 1, & x \in \widetilde{\Omega}_-, \\ 0, & x \in \overline{\widetilde{\Omega}}_+ \cup \partial\Omega. \end{cases}$$

Define $u = g_+ - g_-$ and $v = -\operatorname{div}(A\nabla u)$ in the weak sense. Then $u = Kv$ and u satisfies the conditions in (13) for Kv, i.e., v is the required source element in $L^2(\Omega)$. Hence the existence of sets $\widetilde{\Omega}_+$ and $\widetilde{\Omega}_-$ with properties (14) is actually the sufficient condition for a minimum norm solution in the case of the elliptic problem (6) to fulfill the source condition (13).

Note that sparse minimum norm solutions, i.e.,

$$\mu^\dagger = \sum_{i=1}^{N} a_j \delta_{x_j}, \quad a_j \in \mathbb{R},\ x_j \in \Omega,$$

always satisfy the source condition (13).

4 A numerical minimization algorithm

The following section aims at proposing a numerical algorithm for the iterative minimization of the Tikhonov functional (3) and to derive its convergence. From the numerical analysis point of view, this seems to be challenging as we are dealing with algorithms for the non-reflexive space $\mathcal{M}(\Omega, \mathbb{R}^m)$

which moreover does not possess any smoothness, convexity or other properties useful for studying convergence in Banach spaces [47, 34]. On the other hand, Radon measures are well-studied objects and many notions of convergence are available (see, for instance, [23, 26]) for which one can hope that it holds for a numerical algorithm.

We propose an extension of the well-known iterative soft-thresholding algorithm for the solution of inverse problems in ℓ^2 with (weighted) ℓ^1-penalization [14], i.e.,

$$\min_{u \in \ell^2} \frac{\|Au - f\|_2^2}{2} + \alpha \sum_{k=1}^{\infty} |u_k|.$$

The algorithm proceeds as follows: Take a stepsize $0 < s < 2/\|A\|^2$ and iterate

$$\begin{cases} u^0 \in \ell^2 \\ u^{n+1} = \mathbf{S}_{\alpha s}\big(u^n - sA^*(Au^n - f)\big) \end{cases}$$

where $\mathbf{S}_{\alpha s}$ denotes componentwise soft-thresholding defined by

$$\mathbf{S}_{\alpha s}(v)_k = \operatorname{sgn}(v_k)[|v_k| - \alpha s]_+ = \tfrac{v_k}{|v_k|} \max\,\{0, |v_k| - \alpha s\}.$$

This algorithm is known to converge in the sequence-space setting, in many cases at a q-linear rate [14, 12, 7]. However, this algorithm has no natural generalization to Radon measures, having its main reason in the failure of point evaluation for a $\mu \in \mathcal{M}(\Omega, \mathbb{R}^m)$ in contrast to some $u \in \ell^2$.

This defect can, nevertheless, easily be overcome by going to the discrete setting as follows. Consider again problem (3) in the situation of Proposition 1. We now assume that the sought measure μ consists of finitely many delta peaks

$$\mu = \sum_{x_k \in \Omega_h} u_k \delta_{x_k}$$

where $\Omega_h \subset \Omega$ such that $\#\{\Omega_h\} = N < \infty$ and each $u_k \in \mathbb{R}^m$. Therefore, writing

$$u = (u_1, \ldots, u_N) \quad , \quad Au = \sum_{x_k \in \Omega_h} u_k K^* \delta_{x_k}$$

gives a linear and continuous operator between finite-dimensional vector spaces. Since the ansatz measures are sums of delta peaks, the discretized problem then reads as

$$\min_{u \in \mathbb{R}^{Nm}} \frac{\|Au - f\|_2^2}{2} + \alpha \sum_{k=1}^{N} |u_k|.$$

The iterative thresholding procedure can now be performed on u and is known to converge (note that u_k is penalized by a vector norm. The above corresponds to joint sparsity constraints for which the analysis is, for instance, carried out in [24]). Indeed, this approach is often taken in practice when an $L^1(\Omega, \mathbb{R}^m)$ or $\mathcal{M}(\Omega, \mathbb{R}^m)$ regularization term is considered [43]. Our approach, however, stays in the continuous domain and incorporates a way of adding and removing delta peaks at arbitrary points of Ω.

Throughout this section, let Ω be a separable locally compact metric space, H a Hilbert space, $K : H \to \mathcal{C}_0(\Omega, \mathbb{R}^m)$ a linear and continuous mapping, $f \in H$ given data and $\alpha > 0$ a fixed regularization parameter. Subsequently, we investigate an algorithm for the minimization of the Tikhonov functional T_α in (3). In order to formulate the algorithm, we need some prerequisites. First, define, for a $t \geq 0$,

$$\varphi(t) = \begin{cases} \alpha t & \text{if } t \leq M_0, \\ \frac{\alpha}{2M_0}(t^2 + M_0^2) & \text{if } t > M_0, \end{cases} \qquad M_0 = \frac{\|f\|^2}{2\alpha}.$$

Furthermore, for $\sigma, \tau \in \mathbb{R}^m$ we associate the set-valued sign mappings

$$\begin{aligned} \operatorname{Sgn}(\tau) &= \{\sigma \in \mathbb{R}^m \ : \ |\sigma|_* \leq 1,\ \sigma \cdot \tau = |\tau|\}, \\ \operatorname{Sgn}_*(\sigma) &= \{\tau \in \mathbb{R}^m \ : \ |\tau| \leq 1,\ \sigma \cdot \tau = |\sigma|_*\} \end{aligned}$$

as well as $\operatorname{sgn}(\tau) = \sigma$ such that $\sigma = \operatorname{argmin}_{\tilde{\sigma} \in \operatorname{Sgn}(\tau)} |\tilde{\sigma}|_2$ and the analogous construction for sgn_*. Note that in case of $|\cdot|$ being the Euclidean norm, we have $\operatorname{Sgn}(\sigma) = \{\sigma/|\sigma|\}$ for $\sigma \neq 0$ and $\operatorname{Sgn}(0) = \{|\sigma| \leq 1\}$. In particular, in the dimension one it holds that $\operatorname{Sgn}(s) = 1$ for $s > 0$, $\operatorname{Sgn}(s) = -1$ for $s < 0$ and $\operatorname{Sgn}(0) = [-1, 1]$. To give another example, for $|\tau| = \sum_{i=1}^m |\tau_i|$ it turns out that $\sigma \in \operatorname{Sgn}(\tau)$ if and only if $\sigma_i \in \operatorname{Sgn}(\tau_i)$.

We moreover need the soft-thresholding operation associated with $\beta > 0$ and $w \in \mathbb{R}^m$

$$\mathcal{S}_\beta(w) = \beta(I - P_*)\Big(\frac{w}{\beta}\Big)$$

with P_* being the Euclidean projection on dual unit ball $\{|w|_* \leq 1\}$. For the Euclidean norm $|\cdot| = |\cdot|_2$ on $\mathbb{R}^m$, this corresponds to $\mathcal{S}_\alpha(w) = \operatorname{sgn}(w)[|w| - \alpha]_+$, for instance. Finally, denote by $C_{|\cdot|}$ a constant for which $|w| \leq C_{|\cdot|}|w|_2$ holds for all $w \in \mathbb{R}^m$.

With these prerequisites, the proposed minimization algorithm reads as follows:

Algorithm 1.

1. Initialize $\mu^0 = 0$, $0 < s_* < 2/\|K^*\|^2$ and $n = 0$.

2. For a given $\mu^n = \sum_{j=1}^{J_n} u_j^n \delta_{x_j^n}$ with $u_j^n \neq 0$ and x_j^n pairwise disjoint, compute

$$w^n = -K(K^*\mu^n - f)$$

and determine an $x_*^n \in \Omega$ such that $|w^n(x_*^n)|_* = \|w^n\|_\infty$.

3. Set $\nu^n = v^n \delta_{x_*^n}$ with
$$v^n = \begin{cases} 0 & \text{if } |w^n(x_*^n)|_* \leq \alpha, \\ \alpha^{-1} M_0 \|w^n\|_\infty \operatorname{sgn}_*\big(w^n(x_*^n)\big) & \text{if } |w^n(x_*^n)|_* > \alpha. \end{cases}$$
4. Evaluate, with division by zero resulting in ∞,
$$s_n = \min \Big\{1, \frac{\alpha \sum_{j=1}^{J_n} |u_j^n| - \varphi\big(|v^n|\big) - \langle \mu^n - \nu^n, w^n \rangle}{\|K^*(\mu^n - \nu^n)\|^2} \Big\},$$
stop if $s_n = 0$ with μ^n being a solution. Otherwise, set
$$\mu^{n+1/2} = \mu^n + s_n(\nu^n - \mu^n) = \sum_{j=1}^{J_n+1} u_j^{n+1/2} \delta_{x_j^{n+1}}.$$
5. Set $s_{n+1/2} = s_* / \big(C_{|\cdot|}^2 (J_n + 1)\big)$ and compute
$$u_j^{n+1} = \mathcal{S}_{\alpha s_{n+1/2}} \Big(u_j^n - s_{n+1/2} \big(K(K^* \mu^{n+1/2} - f)\big)(x_j^{n+1}) \Big)$$
for $j = 1, \ldots, J_n + 1$.
6. Remove the x_j^{n+1} for which $u_j^{n+1} = 0$ such that the next iterate becomes
$$\mu^{n+1} = \sum_{j=1}^{J_{n+1}} u_j^{n+1} \delta_{x_j^{n+1}}$$
with $u_j^{n+1} \neq 0$ for $j = 1, \ldots, J_{n+1}$. Increase the counter n and continue with Step 2.

It will turn out that this method is an instance of a combination between a generalized conditional gradient method and the well-known iterative soft-thresholding method (or, more generally, a generalized gradient projection method). We will first establish the connection to the generalized conditional gradient method for which one step is performed in Steps 2–4 in the algorithm.

Proposition 7. *For a $w \in \mathcal{C}_0(\Omega, \mathbb{R}^m)$ a minimizer of*
$$\min_{\nu \in \mathcal{M}(\Omega, \mathbb{R}^m)} -\langle \nu, w \rangle + \varphi(\|\nu\|_{\mathcal{M}}) \tag{15}$$
is given by
$$\nu = \begin{cases} 0 & \textit{if } \|w\|_\infty \leq \alpha, \\ \alpha^{-1} M_0 \|w\|_\infty \operatorname{sgn}_*\big(w(x)\big) \delta_x & \textit{otherwise} \end{cases} \tag{16}$$
with $x \in \Omega$ such that $|w(x)|_* = \|w\|_\infty$.

Moreover, the set-valued operator taking w to all solutions ν of (15) *is bounded on bounded sets.*

Proof. Note that since φ is convex on $[0,\infty[$, a minimizer ν of (15) can be characterized by $j^{**}w \in \partial(\varphi \circ \|\cdot\|_{\mathcal{M}})(\nu)$ where $j^{**} : \mathcal{C}_0(\Omega,\mathbb{R}^m) \to \mathcal{C}_0(\Omega,\mathbb{R}^m)^{**} = \mathcal{M}(\Omega,\mathbb{R}^m)^*$ denoting the canonical injection. From convex analysis it is known that the subgradient of $\nu \mapsto \varphi(\|\nu\|_{\mathcal{M}})$ is given by all $q \in \mathcal{M}(\Omega,\mathbb{R}^m)^*$ with

$$\langle q, \nu\rangle = \|q\|_{\mathcal{M}^*}\|\nu\|_{\mathcal{M}} \quad \text{and} \quad \|q\|_{\mathcal{M}^*} \in \partial\varphi(\|\nu\|_{\mathcal{M}}).$$

In the particular case of $w \in \mathcal{C}_0(\Omega,\mathbb{R}^m)$, this reads as

$$\langle \nu, w\rangle = \|\nu\|_{\mathcal{M}}\|w\|_\infty \quad \text{and} \quad \|w\|_\infty \begin{cases} \leq \alpha & \text{if } \nu = 0 \\ = \alpha & \text{if } 0 < \|\nu\|_{\mathcal{M}} \leq M_0, \\ = \frac{\alpha\|\nu\|_{\mathcal{M}}}{M_0} & \text{else.} \end{cases} \tag{17}$$

Hence, $\nu = 0$ is a minimizer if $\|w\|_\infty \leq \alpha$. In the case $\|w\|_\infty > \alpha$, choose a $x \in \Omega$ such that $|w(x)|_* = \|w\|_\infty$ which exists by continuity. Hence, $\nu = \alpha^{-1}M_0\|w\|_\infty \operatorname{sgn}_*\big(w(x)\big)\delta_x$ satisfies $\|\nu\|_{\mathcal{M}} > M_0$ and the definition of the sign operator yields

$$\begin{aligned} \langle \nu, w\rangle &= \alpha^{-1}M_0\|w\|_\infty \operatorname{sgn}_*\big(w(x)\big) \cdot w(x) \\ &= \alpha^{-1}M_0\|w\|_\infty\|w\|_\infty = \|\nu\|_{\mathcal{M}}\|w\|_\infty \end{aligned}$$

as well as $\frac{\alpha\|\nu\|_{\mathcal{M}}}{M_0} = \|w\|_\infty$. Consequently, ν is a minimizer of (15).

Eventually, let $w \in \mathcal{C}_0(\Omega,\mathbb{R}^m)$ and $\nu \in \mathcal{M}(\Omega,\mathbb{R}^m)$ be a minimizer of (15). According to (17), each solution satisfies $\|\nu\|_{\mathcal{M}} \leq M_0$ or $\|\nu\|_{\mathcal{M}} = \frac{M_0\|w\|_\infty}{\alpha}$ implying that the operator taking w to all possible solutions ν is bounded on bounded sets. □

Next, we establish the relation of this conditional gradient step to the problem of minimizing the Tikhonov functional (3), noting in particular that we can replace the term $\alpha\|\cdot\|_{\mathcal{M}}$ by $\varphi\circ\|\cdot\|_{\mathcal{M}}$ without changing the solutions.

Lemma 1.

(i) The minimizers of T_α are exactly the minimizers of

$$\tilde{T}_\alpha(\mu) = \frac{\|K^*\mu - f\|^2}{2} + \varphi\big(\|\mu\|_{\mathcal{M}}\big).$$

(ii) For $\mu \in \mathcal{M}(\Omega,\mathbb{R}^m)$, $w = -K(K^*\mu - f)$ and ν a solution of (15), it holds that

$$G(\mu,\nu) := \alpha\|\mu\|_{\mathcal{M}} - \varphi\big(\|\nu\|_{\mathcal{M}}\big) - \langle \mu - \nu, w\rangle$$

satisfies $G(\mu,\nu) \geq T_\alpha(\mu) - T_\alpha(\mu^*)$ for any minimizer μ^* of T_α.

Proof. (i) Let μ^* be a minimizer of T_α. Necessarily,

$$\alpha\|\mu^*\|_{\mathcal{M}} \leq T_\alpha(\mu^*) \leq T_\alpha(0) = \frac{\|f\|^2}{2} \quad\implies\quad \|\mu^*\|_{\mathcal{M}} \leq \frac{\|f\|^2}{2\alpha} = M_0.$$

On the other hand, if μ^* minimizes T_α over $\{\|\mu\|_{\mathcal{M}} \leq M_0\}$, then it is also a minimizer of T_α on the whole $\mathcal{M}(\Omega, \mathbb{R}^m)$: otherwise there would be a minimizer μ^{**} of T_α with $T_\alpha(\mu^{**}) < T_\alpha(\mu^*)$. But, $\|\mu^{**}\|_{\mathcal{M}} \leq M_0$, hence μ^* was not a minimizer of T_α on $\{\|\mu\|_{\mathcal{M}} \leq M_0\}$.

Analogously, one sees that the minimizers of $\tilde{T}_\alpha$ satisfy $\|\mu\|_{\mathcal{M}} \leq M_0$ and coincide with the minimizers of $\tilde{T}_\alpha$ on $\{\|\mu\|_{\mathcal{M}} \leq M_0\}$. But, $T_\alpha = \tilde{T}_\alpha$ on $\{\|\mu\|_{\mathcal{M}} \leq M_0\}$, hence the minimizers of T_α and $\tilde{T}_\alpha$ are the same.

(ii) One easily checks that for each $\mu^* \in \mathcal{M}(\Omega, \mathbb{R}^m)$

$$-\langle \mu - \mu^*, w\rangle = \frac{\|K^*\mu - f\|^2}{2} - \frac{\|K^*\mu^* - f\|^2}{2} + \frac{\|K^*(\mu^* - \mu)\|^2}{2}$$

by bilinear computations. According to (i), we have for any minimizer μ^* of T_α that $\varphi(\|\mu^*\|_{\mathcal{M}}) = \alpha\|\mu^*\|_{\mathcal{M}}$ and one easily verifies that $\varphi(\|\mu\|_{\mathcal{M}}) \geq \alpha\|\mu\|_{\mathcal{M}}$ for all μ. This leads, together with the fact that ν solves (15), to

$$\begin{aligned}
G(\mu, \nu) &\geq -\langle \mu, w\rangle + \alpha\|\mu\|_{\mathcal{M}} - \Big(\min_{\bar{\mu}\in\mathcal{M}(\Omega,\mathbb{R}^m)} -\langle \bar{\mu}, w\rangle + \varphi\big(\|\bar{\mu}\|_{\mathcal{M}}\big)\Big) \\
&\geq -\langle \mu - \mu^*, w\rangle + \alpha\|\mu\|_{\mathcal{M}} - \alpha\|\mu^*\|_{\mathcal{M}} \\
&\geq \frac{\|K^*\mu - f\|^2}{2} - \frac{\|K^*\mu^* - f\|^2}{2} + \alpha\|\mu\|_{\mathcal{M}} - \alpha\|\mu^*\|_{\mathcal{M}} \\
&= T_\alpha(\mu) - T_\alpha(\mu^*).
\end{aligned}$$

Hence the inequality holds as stated. □

Knowing that the choice of the delta peak in Step 3 of Algorithm 1 corresponds to the solution of the partially linearized problem (15), we further can derive a descent estimate for the functional.

Proposition 8. *There exists a $c > 0$ such that for each $\mu \in \mathcal{M}(\Omega, \mathbb{R}^m)$ with $T_\alpha(\mu) \leq \|f\|^2/2$ and $w = -K(K^*\mu - f)$, the measure $\bar{\mu} = \mu + s(\nu - \mu)$ with ν being a solution of* (15) *and*

$$s = \min\Big\{1, \frac{\alpha\|\mu\|_{\mathcal{M}} - \varphi\big(\|\nu\|_{\mathcal{M}}\big) - \langle \mu - \nu, w\rangle}{\|K^*(\mu - \nu)\|^2}\Big\}$$

satisfies, for any minimizing argument μ^* of T_α,

$$r(\bar{\mu}) - r(\mu) \leq -cr(\mu)^2, \quad r(\mu) = T_\alpha(\mu) - T_\alpha(\mu^*).$$

Proof. Observe again that $T_\alpha(\mu) \le \|f\|^2/2 = T_\alpha(0)$ implies $\|\mu\|_{\mathcal{M}} \le M_0$ and that easy bilinear computations show

$$\frac{\|K^*\bar{\mu} - f\|^2}{2} - \frac{\|K^*\mu - f\|^2}{2} = s\langle K^*\mu - f,\, K^*(\nu - \mu)\rangle + \frac{s^2\|K^*(\nu-\mu)\|^2}{2}.$$

Since $\varphi \circ \|\cdot\|_{\mathcal{M}}$ is convex, we also have

$$\varphi\big(\|\bar{\mu}\|_{\mathcal{M}}\big) - \varphi\big(\|\mu\|_{\mathcal{M}}\big) \le s\big(\varphi\big(\|\nu\|_{\mathcal{M}}\big) - \varphi\big(\|\mu\|_{\mathcal{M}}\big)\big).$$

Together, noting that $\varphi(\|\mu\|_{\mathcal{M}}) = \alpha\|\mu\|_{\mathcal{M}}$, it follows that

$$\begin{aligned}\tilde{T}_\alpha(\bar{\mu}) - \tilde{T}_\alpha(\mu) &\le s\big(\langle \mu - \nu,\, w\rangle + \varphi\big(\|\nu\|_{\mathcal{M}}\big) - \alpha\|\mu\|_{\mathcal{M}}\big) + \frac{s^2\|K^*(\nu-\mu)\|^2}{2}\\ &= -sG(\mu,\nu) + \frac{s^2\|K^*(\nu-\mu)\|^2}{2}. \qquad (18)\end{aligned}$$

We now distinguish two cases: First, suppose $\|K^*(\nu-\mu)\|^2 \le G(\mu,\nu)$ leading to $s = 1$ and, consequently, with the help of Lemma 1, to

$$\tilde{T}_\alpha(\bar{\mu}) - \tilde{T}_\alpha(\mu) \le -G(\mu,\nu) + \frac{G(\mu,\nu)}{2} = -\frac{G(\mu,\nu)}{2} \le -\frac{r(\mu)}{2}.$$

With

$$r(\mu) = T_\alpha(\mu) - T_\alpha(\mu^*) \le T_\alpha(\mu) \le \frac{\big(\|K^*\|M_0 + \|f\|\big)^2}{2} + \alpha M_0 \le C_1,$$

such that $C_1 > 0$, the above becomes

$$\tilde{T}_\alpha(\bar{\mu}) - \tilde{T}_\alpha(\mu) \le -\frac{r(\mu)^2}{2C_1}.$$

We like to prove this kind of estimate also if $\|K^*(\nu-\mu)\|^2 > G(\mu,\nu)$. In this case, $s = G(\mu,\nu)/\|K^*(\nu-\mu)\|^2 < 1$, so it plugging into (18), we arrive at

$$\begin{aligned}\tilde{T}_\alpha(\bar{\mu}) - \tilde{T}_\alpha(\mu) &\le -\frac{G(\mu,\nu)^2}{\|K^*(\nu-\mu)\|^2} + \frac{G(\mu,\nu)^2}{2\|K^*(\nu-\mu)\|^2} = -\frac{G(\mu,\nu)^2}{2\|K^*(\nu-\mu)\|^2}\\ &\le -\frac{r(\mu)^2}{2\|K^*(\nu-\mu)\|^2}.\end{aligned}$$

Since $w = -K(K^*\mu - f)$ is bounded if μ is bounded and ν is bounded if w is bounded (by, say $C_2 > 0$, see Proposition 7), there follows $\|K^*(\nu-\mu)\|^2 \le \|K^*\|^2(C_2 + M_0)^2 \le C_3$ for some $C_3 > 0$. Thus

$$\tilde{T}_\alpha(\bar{\mu}) - \tilde{T}_\alpha(\mu) \le -\frac{r(\mu)^2}{2C_3}.$$

In both of the two cases, $\tilde{T}_\alpha(\bar{\mu}) \le \tilde{T}_\alpha(\mu) = T_\alpha(\mu) \le T_\alpha(0)$ and hence, $\|\bar{\mu}\|_{\mathcal{M}} \le M_0$ leads to $\tilde{T}_\alpha(\bar{\mu}) = T_\alpha(\bar{\mu})$. Finally, by choosing $c = \min\,\{1/(2C_1), 1/(2C_3)\} > 0$ the desired estimate $r(\bar{\mu}) - r(\mu) \le -cr(\mu)^2$ is obtained. □

Taking a look at Algorithm 1, we see that ν chosen in Steps 2–3 corresponds to a minimizing element for (15) (see Proposition 7), so Step 4 produces $\mu^{n+1/2} = \bar{\mu}^n$ according to Proposition 8. Hence, there is a $c > 0$ independent of n, such that

$$r(\mu^{n+1/2}) - r(\mu^n) \leq -cr(\mu^n)^2$$

if $T_\alpha(\mu^n) \leq \|f\|^2/2$ for all n. For μ^0 this is indeed true, and it would be for all n, if also $r(\mu^{n+1}) \leq r(\mu^{n+1/2})$ for all n. Showing the latter is the goal of the following proposition.

Proposition 9. *Let $J \geq 1$, $x_1, \ldots, x_J \in \Omega$ be pairwise disjoint, and $B : (\mathbb{R}^m)^J \to \mathcal{M}(\Omega, \mathbb{R}^m)$ be defined by $Bu = \sum_{j=1}^J u_j \delta_{x_j}$. Then, for all $u_1, \ldots, u_J \in \mathbb{R}^m$ and $0 < s < 2/(C_{|\,\cdot\,|}\|K^*\|^2 J)$, it holds that*

$$v_j = \mathcal{S}_{s\alpha}\Big(u_j - s\big(B^*K(K^*Bu - f)\big)_j\Big), \quad j = 1, \ldots, J, \tag{19}$$

satisfies

$$\frac{\|K^*Bv - f\|^2}{2} + \alpha \sum_{j=1}^J |v_j| \leq \frac{\|K^*Bu - f\|^2}{2} + \alpha \sum_{j=1}^J |u_j|. \tag{20}$$

Proof. First note that the norm of the operator B can be estimated as follows: For $u_1, \ldots, u_J \in \mathbb{R}^m$,

$$\|Bu\|_{\mathcal{M}} = \sum_{j=1}^J |u_j| \leq \Big(\sum_{j=1}^J 1\Big)^{1/2} \Big(\sum_{j=1}^J |u_j|^2\Big)^{1/2} \leq \sqrt{J} C_{|\,\cdot\,|} \Big(\sum_{j=1}^J |u_j|_2^2\Big)^{1/2}$$

implying that $\|B\| \leq \sqrt{J} C_{|\,\cdot\,|}$.

It is well-known [14, 24] that the soft-thresholding operator $\eta = \mathcal{S}_\beta(\xi)$ is the solution operator to

$$\min_{\eta \in \mathbb{R}^m} \frac{|\eta - \xi|_2^2}{2} + \beta|\eta|. \tag{21}$$

We nevertheless give a short argumentation for the sake of completeness. It is easily seen, mostly analogous to Proposition 2, that the Fenchel dual problem is equivalent to

$$\min_{\zeta \in \mathbb{R}^m} \frac{|\zeta - \xi|_2^2}{2} + I_{\{|\zeta|_* \leq \beta\}}(\zeta)$$

with the respective solutions satisfying $\eta^* = \xi - \zeta^*$. The above problem is just the projection of ξ onto the ball $\{|\zeta|_* \leq \beta\}$ which can be expressed as $P_{\{|\xi|_* \leq \beta\}}(\xi) = \beta P_*(\xi/\beta)$. Together, it follows that $\eta^* = \beta(I - P_*)(\xi/\beta) = \mathcal{S}_\beta(\xi)$.

Denote by $w_j = \big(-B^*K(K^*Bu - f)\big)_j$. Since $v_j = \mathcal{S}_{s\alpha}(u_j + sw_j)$ is the solution of the problem (21), we can express it with subgradients and use subgradient calculus:

$$0 \in \partial\Big(\frac{|\cdot - u_j - sw_j|_2^2}{2} + \alpha s|\cdot|\Big)(v_j) \quad \Leftrightarrow \quad u_j + sw_j - v_j \in s\alpha\partial|\cdot|(v_j).$$

In terms of the subgradient inequality, the latter just reads as

$$(u_j + sw_j - v_j)\cdot(\eta_j - v_j) \le s\big(\alpha|\eta_j| - \alpha|v_j|\big).$$

Introducing the quantities $G_j = \alpha|u_j| - \alpha|v_j| - w_j\cdot(u_j - v_j)$, setting $\eta_j = u_j$ and rearranging yields

$$|u_j - v_j|_2^2 \le s\big(\alpha|u_j| - \alpha|v_j| + w_j\cdot(u_j - v_j)\big) = sG_j. \tag{22}$$

In particular, $G_j \ge 0$. Next, observe that

$$\frac{\|K^*Bv - f\|^2}{2} - \frac{\|K^*Bu - f\|^2}{2} - \langle K^*Bu - f,\, K^*B(v-u)\rangle = \frac{\|K^*B(v-u)\|^2}{2}.$$

So using $\sum_{j=1}^J w_j\cdot(u_j - v_j) = \langle K^*Bu - f,\, K^*B(v-u)\rangle$, the estimate on $\|B\|$, (22), the choice of s and (22) again, we get

$$\begin{aligned}
&\frac{\|K^*Bv - f\|^2}{2} - \frac{\|K^*Bu - f\|^2}{2} + \alpha\sum_{j=1}^J |v_j| - \alpha\sum_{j=1}^J |u_j| \\
&= \frac{\|K^*B(v-u)\|^2}{2} + \langle K^*Bu - f,\, K^*B(v-u)\rangle + \alpha\Big(\sum_{j=1}^J |v_j| - |u_j|\Big) \\
&= \frac{\|K^*B(v-u)\|^2}{2} - \sum_{j=1}^J \alpha|u_j| - \alpha|v_j| + w_j\cdot(u_j - v_j) \\
&\le \frac{\|K^*\|^2\|B\|^2}{2}\sum_{j=1}^J |v_j - u_j|_2^2 - \sum_{j=1}^J G_j \\
&\le \Big(\frac{JC_{|\cdot|}^2\|K^*\|^2}{2} - \frac{1}{s}\Big)\sum_{j=1}^J |u_j - v_j|_2^2 \le 0.
\end{aligned}$$

The desired statement now immediately follows. □

By examining the notation, it is easy to see that Step 5 of Algorithm 1 corresponds to the thresholding step (19) and that (20) is just $r(\mu^{n+1}) \le r(\mu^{n+1/2})$. Moreover, neglecting delta peaks with zero height as done in Step 6 does not change μ^{n+1}. Thus, we have all the ingredients to prove the subsequential weak* convergence of Algorithm 1 to minimizers of the Tikhonov functional (3).

Theorem 3. *Algorithm 1 produces a minimizing sequence $\{\mu^n\}$ of measures for the functional T_α with rate $r(\mu^n) \leq Cn^{-1}$ for some $C > 0$. Each subsequence of $\{\mu^n\}$ possesses a weak* convergent subsequence whose limit μ^* is a minimizer of T_α. If the minimizer is unique, the whole sequence converges to μ^* in the weak* sense.*

Proof. Let us first prove, by induction, that $\{r(\mu^n)\}$ is a non-increasing sequence, i.e., $r(\mu^n) \leq r(\mu^{n-1}) \leq \ldots \leq r(\mu^0)$ for all n. If this holds for some n, $r(\mu^n) \leq r(\mu^0)$ and consequently, $T_\alpha(\mu^n) \leq T_\alpha(\mu^0) = \|f\|^2/2$, so Proposition 8 can be applied (note again, that $\mu^{n+1/2}$ corresponds to $\bar{\mu}^n$ by virtue of Proposition 7). This leads to $r(\mu^{n+1/2}) \leq r(\mu^n)$ and by virtue of Proposition 9, we also have $r(\mu^{n+1}) \leq r(\mu^n)$. Hence, $\{r(\mu^n)\}$ is non-increasing.

To prove that $\{\mu^n\}$ is a minimizing sequence, it suffices to obtain the rate. Note that Propositions 8 and 9 also yield a $c > 0$ such that

$$r(\mu^{n+1}) - r(\mu^n) \leq -cr(\mu^n)^2$$

for each n. Denoting $r_n = r(\mu^n)$ we employ the following widely-known trick: Estimate

$$\frac{1}{r_{n+1}} - \frac{1}{r_n} = \frac{r_n - r_{n+1}}{r_{n+1} r_n} \geq \frac{cr_n^2}{r_{n+1} r_n} \geq c > 0,$$

sum up

$$\frac{1}{r_n} - \frac{1}{r_0} = \sum_{i=0}^{n-1} \frac{1}{r_{i+1}} - \frac{1}{r_i} \geq \sum_{i=0}^{n-1} c = cn,$$

and rearrange to $r_n \leq (r_0^{-1} + cn)^{-1} \leq Cn^{-1}$ for some $C > 0$.

The rest can be deduced from functional-analytic arguments: Any subsequence of $\{\mu^n\}$ is also a minimizing sequence and hence, another subsequence exists for which there is a weak* limit μ^* which is a minimizer (the latter from the weak* lower semi-continuity, also see Proposition 1). In case of uniqueness of μ^*, the statement follows from the usual subsequence argument. □

Remark 8. It is obvious from the proof that Algorithm 1 does still converge if the soft-thresholding step (Step 5) is left out. We added this step, however, to encourage the removal of peaks from the iterates. This may be helpful in practical implementations, as it is usually more efficient to deal with a lower number of peaks.

Note, moreover, that the algorithm also still converges if Step 5 is replaced with or supplemented by any method which do not increase the functional value. As it is shown in Proposition 9, performing one soft-thresholding iteration belongs to the class of such methods. One can as well repeat this iteration several times and the overall convergence as well as the rate will not be affected provided that the number of repetitions stays bounded.

In this spirit, the soft-thresholding algorithm, may be replaced by suitable application of an algorithm with better asymptotic convergence rate, like, for instance of Nesterov-type [40] like the "fast iterative shrinkage-thresholding algorithm" (FISTA) [3], if sufficiently many iterations are carried out such that the functional value is not increased.

Provided that the predual forward operator K is mapping into the space of continuously differentiable functions, it is moreover possible to insert a step which moves the position of the peaks according to the gradient flow of the discrepancy functional with respect to the peak positions. Concerning a single peak, this idea has also been presented in [38], here we adapt it to our general framework.

Proposition 10. *Suppose that, $\Omega \subset \mathbb{R}^d$ is a non-empty, open subset and $K : H \to \mathcal{C}_0^1(\Omega, \mathbb{R}^m)$ is linear and continuous. Then, for each $\mu = \sum_{j=1}^J u_j \delta_{x_j^0}$ with $u_j \neq 0$ and $x_j^0 \in \Omega$ pairwise disjoint, a $t_0 > 0$ and a solution of the ordinary differential equation*

$$\frac{\partial x_j(t)}{\partial t} = -\nabla\Big(K\Big(K^* \sum_{j=1}^J u_j \delta_{x_j(t)} - f\Big)\Big)\big(x_j(t)\big) \qquad , \qquad x_j(0) = x_j^0 \tag{23}$$

exists for $t \in \,]0, t_0[$. Each solution on $]0, t_0[$ satisfies, with $\mu(t) = \sum_{j=1}^J u_j \delta_{x_j(t)}$,

$$\frac{\partial T_\alpha\big(\mu(t)\big)}{\partial t} \leq 0 \qquad \textit{for all} \qquad t \in \,]0, t_0[.$$

Proof. We first show that the mapping $\Phi : \Omega^J \to \mathbb{R}$ given by

$$\Phi(x) = T_\alpha\big(\mu(x)\big) \qquad , \qquad \mu(x) = \sum_{j=1}^J u_j \delta_{x_j}$$

is continuously differentiable with gradient

$$\big(\nabla \Phi(x)\big)_j = \nabla\Big(K\big(K^* \mu(x) - f\big)\Big)(x_j).$$

For that purpose, let $h = (h_1, \ldots, h_J) \in (\mathbb{R}^m)^J$ and consider, for $0 < \tau < \tau_0$ with τ_0 sufficiently small such that $x + \tau h \in \Omega^J$ and $(x + \tau h)_J$ pairwise distinct. Then, the difference quotient reads as

$$\begin{aligned}
&\frac{1}{\tau}\Big(T_\alpha\big(\mu(x+\tau h)\big) - T_\alpha\big(\mu(x)\big)\Big) \\
&\qquad = \frac{1}{\tau}\Big\langle K^*\big(\mu(x+\tau h) - \mu(x)\big),\ \frac{1}{2}K^*\big(\mu(x+\tau h) + \mu(x)\big) - f\Big\rangle \\
&\qquad = \Big\langle \frac{1}{\tau}\big(\mu(x+\tau h) - \mu(x)\big),\ K\big(K^*\mu(x) - f\big)\Big\rangle \\
&\qquad\quad - \frac{\|K^*\big(\mu(x+\tau h) - \mu(x)\big)\|^2}{2\tau}.
\end{aligned}$$

We examine the convergence of the first term. First, denote, once again, by $w = -K(K^*\mu(x) - f)$, so it follows by the fact that K maps into $\mathcal{C}_0^1(\Omega, \mathbb{R}^m)$,

$$\begin{aligned}
&\lim_{\tau\to 0}\frac{1}{\tau}\langle \mu(x+\tau h) - \mu(x),\ K\big(K^*\mu(x) - f\big)\rangle \\
&\qquad\qquad = \lim_{\tau\to 0} -\sum_{j=1}^{J}\frac{w(x_j + \tau h_j) - w(x_j)}{\tau} = \sum_{j=1}^{J} -(\nabla w)(x_j)\cdot h_j.
\end{aligned}$$

On the other hand, estimate the norm of $\mu(x+\tau h) - \mu(x)$ in $\mathcal{C}_0^1(\Omega, \mathbb{R}^m)^*$ by testing with some $v \in \mathcal{C}_0^1(\Omega, \mathbb{R}^m)$:

$$\begin{aligned}
\langle \mu(x+\tau h) - \mu(x),\ v\rangle &= \sum_{j=1}^{J} v(x_j + \tau h_j) - v(x_j) \\
&\leq |\tau|\sum_{j=1}^{J}\int_0^1 |\nabla v(x_j + s\tau h_j)||h_j|\ \mathrm{d}s \leq C|\tau|\|\nabla v\|_\infty,
\end{aligned}$$

hence $\|\mu(x+\tau h) - \mu(x)\|_{(\mathcal{C}^1)^*} \leq C|\tau|$. The second term thus converges to zero since

$$0 \leq \lim_{\tau\to 0}\frac{\|K^*\big(\mu(x+\tau h) - \mu(x)\big)\|^2}{2\tau} \leq \lim_{\tau\to 0}\frac{C^2\tau^2\|K^*\|^2}{2\tau} = 0.$$

This shows that $x \mapsto T_\alpha\big(\mu(x)\big)$ is Gâteaux-differentiable with the claimed derivative. The Fréchet-differentiability then follows from the continuity of the Gâteaux-derivative with respect to x and the finite-dimensionality of $\Omega^J \subset (\mathbb{R}^d)^J$.

Thus, due to Peano's existence theorem, the ordinary differential equation (23) has a solution on an interval $[0, t_0[$ for some $t_0 > 0$. If $x : [0, t_0[\to \Omega^J$ is a solution, we finally see that, with $\mu(t) = \mu\big(x(t)\big)$, for each $t \in]0, t_0[$

$$\frac{\partial(T_\alpha \circ \mu)}{\partial t}(t) = \Big\langle \frac{\partial x}{\partial t}(t),\ \nabla\Phi\big(\mu(t)\big)\Big\rangle = -\sum_{j=1}^{J}\big|\nabla\Phi\big(x(t)\big)_j\big|^2 \leq 0$$

which completes the proof. □

Consequently, one gets functional descent if one inserts, in Step 6, before returning to Step 2, the solution of (23) with u_j and x_j^0 such that $\mu^{n+1} = \sum_{j=1}^J u_j \delta_{x_j^0}$ and replaces μ^{n+1} by $\sum_{j=1}^J u_j \delta_{x(t)_j}$ for some $t \in]0, t_0[$. In practice, one has to employ numerical schemes for the solution of the ODE. A scheme and the associated step-lengths should, of course, be chosen such that it does not increase the value of the Tikhonov-functional, for instance, one could perform a step of the explicit Euler scheme with a step-length which ensures functional descent.

Finally, we like to mention that one could also employ some trial-and-error peak merging approaches to reduce the functional, for instance, if Ω is a convex subset of a normed space, choosing

$$\mu^n = \sum_{j=1}^J u_j \delta_{x_j} \quad , \quad 1 \le j_0 < j_1 \le J \quad ,$$

$$\mu^{n+1} = \begin{cases} (u_{j_0} + u_{j_1}) \delta_{\frac{1}{2}(x_{j_0} + x_{j_1})} + \sum\limits_{j \neq j_0,\, j \neq j_1} u_j \delta_{x_j} & \text{if } T_\alpha(\mu^{n+1}) \le T_\alpha(\mu^n) \\ \mu^n & \text{else} \end{cases} \tag{24}$$

for some j_0 and j_1, for instance chosen such that $|x_{j_0} - x_{j_1}|_2$ is minimal. Again, this step supplements Algorithm 1 without affecting convergence.

To summarize, following steps could be inserted after Step 5 to improve convergence:

- Multiple steps of the soft-thresholding procedure in Step 5 or suitably many steps of FISTA instead,
- solving the ordinary differential equation (23) for some positive time, if possible,
- trying to merge some peaks according to the trial-and-error method (24), if possible.

5 Numerical computations

To show the performance of Algorithm 1, we apply it for a one-dimensional version of the deconvolution problem presented in Example 1. We assume that the convolution kernel k is a cubic spline with respect to a fine uniform grid of size $h_\mathrm{f} > 0$, i.e.,

$$k = \sum_{j=-L+2}^{L-2} k_j \psi_j, \qquad \psi_j(x) = B_0\Big(\frac{x}{h_\mathrm{f}} - j\Big),$$

for some $L \geq 2$ and with the standard cubic box spline

$$B_0(x) = \begin{cases} 0 & \text{for } x < -2, \\ (x+2)^3 & \text{for } -2 \leq x < -1, \\ -3(x+1)^3 + 3(x+1)^2 + 3(x+1) + 1 & \text{for } -1 \leq x < 0, \\ 3(x-1)^3 + 3(x-1)^2 - 3(x-1) + 1 & \text{for } 0 \leq x < 1, \\ -(x-2)^3 & \text{for } 1 \leq x < 2, \\ 0 & \text{for } 2 \leq x. \end{cases}$$

On a possibly coarser grid with size $h_\mathrm{c} = mh_\mathrm{f}$, m being a positive integer, we define the operator K for a $v = (v_0, \ldots, v_{N-1}) \in \mathbb{R}^N$ by

$$(Kv)(x) = \sum_{\kappa=0}^{N-1} v_\kappa k(x - \kappa h_\mathrm{c}) \tag{25}$$

which corresponds to a discrete adjoint convolution of k with the data given in the points $0, h_\mathrm{c}, \ldots, (N-1)h_\mathrm{c}$. Equipping $\mathbb{R}^N$ with the standard scalar product and letting $\Omega = \,]-Lh_\mathrm{f}, (N-1)h_\mathrm{c} + Lh_\mathrm{f}[$, one sees that $K : \mathbb{R}^N \to \mathcal{C}_0(\Omega)$ is linear and continuous and hence, satisfies the requirements of Proposition 1 and in particular of Section 4. It is moreover easy to see that the adjoint corresponds to

$$(K^*\mu)_\kappa = (k * \mu)(\kappa h_\mathrm{c}), \qquad \kappa = 0, \ldots, N-1,$$

i.e., is the convolution of a measure with k evaluated at the grid points with respect to h_c. Hence, with data $f = (f_0, \ldots, f_{N-1})$ defined in these points, the Tikhonov functional to minimize reads as

$$T_\alpha(\mu) = \frac{1}{2}\Big(\sum_{\kappa=0}^{N-1} \big((k * \mu)(\kappa h_\mathrm{c}) - f_\kappa\big)^2\Big) + \alpha \|\mu\|_\mathcal{M}.$$

Note that we still like to minimize over the space $\mathcal{M}(\Omega)$ which we do not discretize.

Algorithm 1 can actually be performed in this semi-continuous setting. For μ consisting of finitely many delta peaks, $K^*\mu$ corresponds to finite point evaluations of k as well as computing finite linear combinations. The function $w = -K(K^*\mu - f)$, however, is defined on a continuum. However, the only thing we have to compute is a point in which $|w|$ becomes maximal, for which it suffices to know the critical points of w. These can be easily obtained since Kv, $v = (v_0, \ldots, v_{N-1}) \in \mathbb{R}^N$, is a cubic spline with respect to the grid h_f (as h_c is a integer multiple of h_f, see (25)):

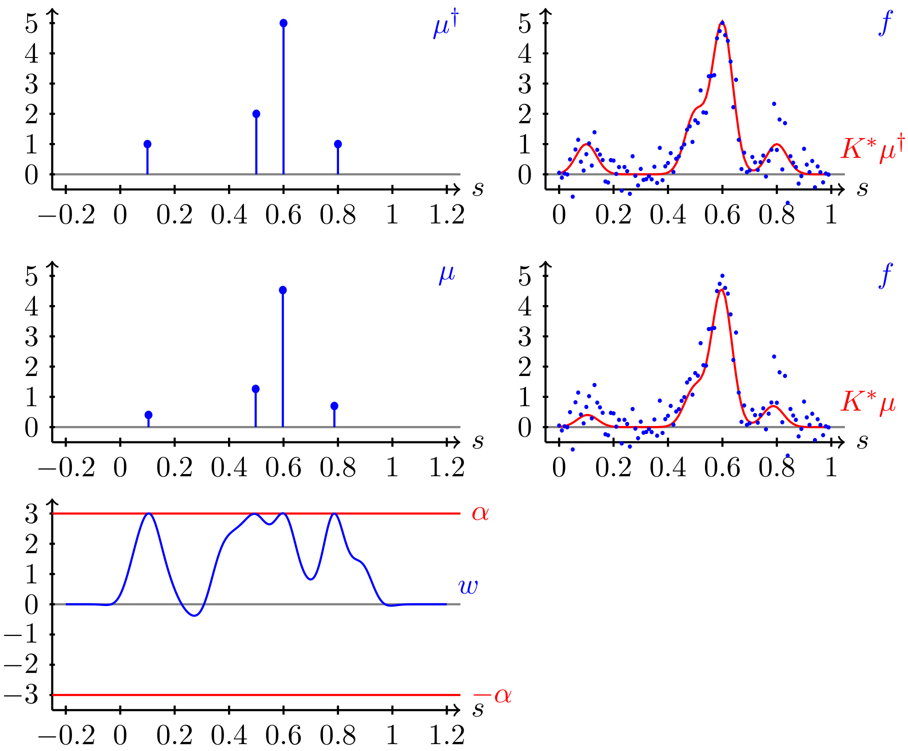
$$Kv = \sum_{\kappa=0}^{N-1} v_\kappa \sum_{j=-L+2}^{L-2} k_j \psi_j(\cdot - \kappa m h_\mathrm{f}) = \sum_{\kappa=0}^{N-1} \sum_{j=-L+2}^{L-2} v_\kappa k_j \psi_{j+\kappa m}.$$

Hence, to compute all critical point of some w, we solve $w'(x) = 0$ on each suitable subinterval $[jh_{\mathrm{f}}, (j+1)h_{\mathrm{f}}]$ which corresponds to solving a quadratic equation (which usually can be done up to machine precision). By determining the maximizer of $|w|$ in these points, we are able to compute x such that $|w(x)| = \|w\|_\infty$. Steps 3–4 of Algorithm 1 can again be evaluated directly and Step 5 only requires finite point evaluation for some $K(K^*\mu - f)$. In summary, this shows that one is able to perform the successive peak insertion and thresholding method in a semi-continuous setup.

We implemented and tested Algorithm 1 for the deconvolution problem. Taking advantage of the special structure of the convolution operator, we also added the ODE-solving step (23) as well as the peak-merging step (24) to improve the performance. The outcome of the method for a Gaussian convolution kernel is depicted in the Figures 1 and 2. As it can be seen, the algorithm is able to reconstruct the sought peaks almost perfectly (with one additional peak introduced in the second example), even in the presence of heavy noise. One has to mention, however, that the regularization results in a decrease of peak size, which is also known to occur in the discrete setting. Moreover, we remark that without the additional steps (23) and (24), the results show significantly more additional peaks. This is closely related to the type of convergence we established for $\{\mu^n\}$: Since it only converges in the weak*-sense, it may happen, for instance, that an indefinite number of delta peaks approximates a single delta peak. Hence, in general, a suitable postprocessing may be necessary which "sparsifies" the outcome of the algorithm (if the aim is to obtain sparse solutions). Nevertheless, the results show that is is possible to minimize, with the help of Algorithm 1, the Tikhonov functional in the semi-discrete setting with continuous solutions μ and discrete data f.

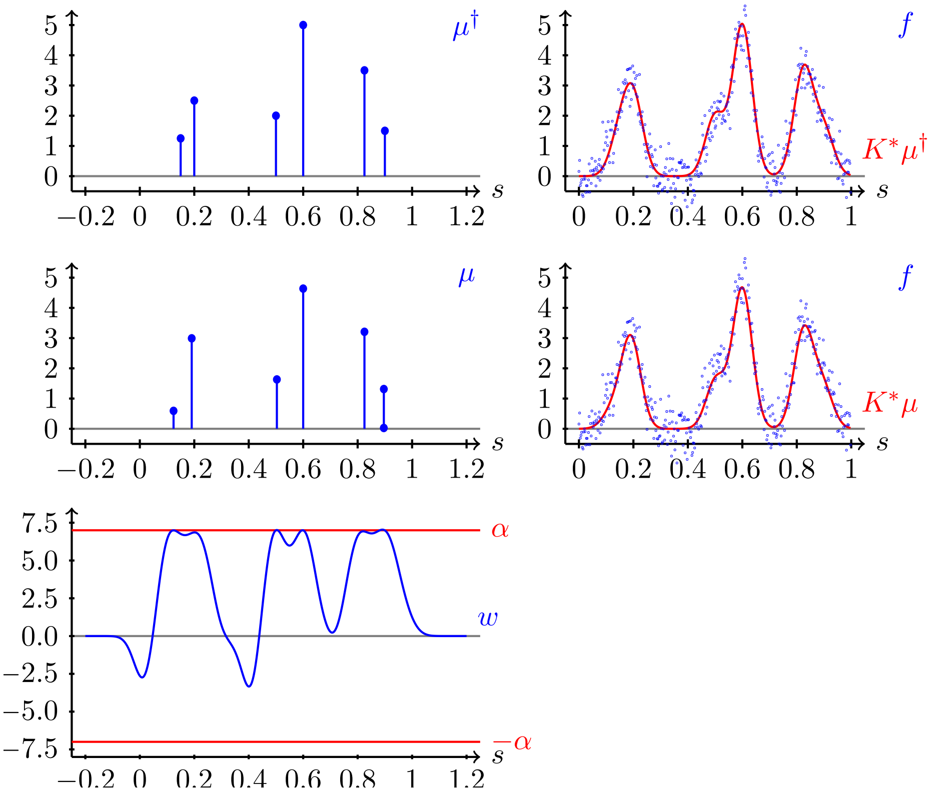
6 Summary and conclusions

Tikhonov regularization for vector-valued finite Radon measures as the solutions for ill-posed linear inverse problems can be put into a proper functional-analytic framework which ensures existence, uniqueness (where applicable) and desireable regularization properties. This framework moreover leads to a flexible abstract numerical algorithm for which descent properties with some rate as well as (subsequential) weak*-convergence can be proven. As it produces delta-peak iterates, the proposed algorithm is well-suited for the recovery of sparse solutions. Furthermore, its structure allows computations directly on the space of Radon measures with only the data space being discretized. However, as it converges only in the weak* sense, it might produce an indefinite number of delta peaks, in particular, if the solution is non-sparse. Nevertheless, numerical computations confirm the applicability of the method and show that it is possible to perform accurate sparse



$\mu^\dagger = \sum_i u_i^\dagger \delta_{x_i^\dagger}$	$x_i^\dagger$	$u_i^\dagger$	$\mu = \sum_i u_i \delta_{x_i}$	x_i	u_i
	0.1000	1.0000		0.1039	0.4017
	0.5000	2.0000		0.4977	1.2623
	0.6000	5.0000		0.5975	4.5303
	0.8000	1.0000		0.7868	0.6979

Figure 1: Numerical deconvolution example for $N = 100$ data points and $h_c = h_f = 0.01$. Upper row: the exact solution $\mu^\dagger$ consisting of four delta peaks (left), its image $K^*\mu^\dagger$ under convolution with a Gaussian kernel with variance $\sigma^2 = 0.05$ and given data f corrupted with Gaussian noise of variance $\sigma^2 = \frac{1}{4}$ (right). Second row: the numerically obtained approximation μ of the minimizer of the associated Tikhonov functional with $\alpha = 3$ (left), its image under convolution $K^*\mu$ compared to the data f (right). Third row: the corresponding $w = -K(K^*\mu - f)$ which, according to the optimality conditions admits α exactly in the peak positions. Bottom row: a table showing the numerical values of $\mu^\dagger$ and μ for comparison.



$\mu^\dagger = \sum_i u_i^\dagger \delta_{x_i^\dagger}$	$x_i^\dagger$	$u_i^\dagger$	$\mu = \sum_i u_i \delta_{x_i}$	x_i	u_i
	0.1500	1.2500		0.1239	0.5926
	0.2000	2.5000		0.1906	2.9924
	0.5000	2.0000		0.5034	1.6321
	0.6000	5.0000		0.6000	4.6377
	0.8250	3.5000		0.8250	3.2083
	0.9000	1.5000		{ 0.8961	0.0265
				0.8962	1.3143

Figure 2: Numerical deconvolution example for $N = 400$ data points and $h_c = h_f = 0.0025$. Upper row: The exact solution $\mu^\dagger$ consisting of six delta peaks (left), the exact and noisy data $K^*\mu^\dagger$ and f with the same parameters as in Figure 1 (right). Second row and third row: the numerical results with $\alpha = 7$ (also see Figure 1). Bottom row: The tabulated numerical values of $\mu^\dagger$ and μ. The solution μ introduced an additional small peak at 0.8961.

deconvolution even in the presence of considerable noise.

References

[1] R. A. Adams and J. J. F. Fournier, *Sobolev spaces*, Academic Press, 2003.

[2] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford University Press, Inc., 2000.

[3] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[4] T. Bonesky, K. S. Kazimierski, P. Maass, F. Schöpfer, and T. Schuster, *Minimization of Tikhonov functionals in Banach spaces*, Abstract and Applied Analysis, 2008 (2008), p. 192679.

[5] K. Bredies, T. Alexandrov, J. Decker, D. A. Lorenz, and H. Thiele, *Sparse deconvolution for peak picking and ion charge estimation in mass spectrometry*, in Progress in Industrial Mathematics at ECMI 2008, H.-G. Bock et al., eds., vol. 15 of Mathematics in Industry, Springer Berlin Heidelberg, 2010, pp. 287–292.

[6] K. Bredies and D. A. Lorenz, *Iterated hard shrinkage for minimization problems with sparsity constraints*, SIAM Journal on Scientific Computing, 30 (2008), pp. 657–683.

[7] K. Bredies and D. A. Lorenz, *Linear convergence of iterative soft-thresholding*, J. Fourier Anal. Appl., 14 (2008), pp. 813–837.

[8] K. Bredies, D. A. Lorenz, and P. Maass, *A generalized conditional gradient method and its connection to an iterative shrinkage method*, Computional Optimization and Applications, 42 (2009), pp. 173–193.

[9] M. Burger and S. Osher, *Convergence rates of convex variational regularization*, Inverse Problems, 20 (2004), pp. 1411–1421.

[10] E. J. Candès, J. K. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Communications on Pure and Applied Mathematics, 59 (2006), pp. 1207–1223.

[11] C. Clason and K. Kunisch, *A duality-based approach to elliptic control problems in non-reflexive Banach spaces*, ESAIM Control Optim. Calc. Var., (2010). doi:10.1051/cocv/2010003.

[12] P. L. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., 4 (2005), pp. 1168–1200.

[13] J. B. Conway, *A course in functional analysis*, vol. 96 of Graduate Texts in Mathematics, Springer, New York, second ed., 1990.

[14] I. Daubechies, M. Defrise, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math., 57 (2004), pp. 1413–1457.

[15] D. L. Donoho, *Compressed sensing*, IEEE Transactions on Information Theory, 52 (2006), pp. 1289–1306.

[16] D. L. Donoho, M. Elad, and V. N. Temlyakov, *Stable recovery of sparse overcomplete representations in the presence of noise*, IEEE Transactions on Information Theory, 52 (2006), pp. 6–18.

[17] C. Dossal and S. Mallat, *Sparse spike deconvolution with minimum scale*, in Proceedings of SPARS'05, November 2005.

[18] N. Dunford and J. T. Schwartz, *Linear Operators. I. General Theory*, With the assistance of W. G. Bade and R. G. Bartle. Pure and Applied Mathematics, Vol. 7, Interscience Publishers, Inc., New York, 1958.

[19] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, *Least angle regression*, The Annals of Statistics, 32 (2004), pp. 407–499.

[20] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publishers, 1996.

[21] H. W. Engl and G. Landl, *Convergence rates for maximum entropy regularization*, SIAM J. Numer. Anal., 30 (1993), pp. 1509–1536.

[22] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, *Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems*, IEEE Journal of Selected Topics in Signal Processing, 1 (2007), pp. 586–597.

[23] I. Fonseca and G. Leoni, *Modern methods in the calculus of variations: L^p spaces*, Springer, 2007.

[24] M. Fornasier and H. Rauhut, *Recovery algorithms for vector valued data with joint sparsity constraints*, SIAM J. Numer. Anal., 46 (2008), pp. 577–613.

[25] J.-J. Fuchs, *On sparse representations in arbitrary redundant bases*, IEEE Transactions on Information Theory, 50 (2004), pp. 1341–1344.

[26] A. L. Gibbs and F. E. Su, *On choosing and bounding probability metrics*, International Statistical Review, 70 (2002), pp. 419–435.

[27] M. Grasmair, M. Haltmeier, and O. Scherzer, *Sparse regularization with ℓ^q penalty term*, Inverse Problems, 24 (2008), p. 055020.

[28] R. Griesse and D. A. Lorenz, *A semismooth Newton method for Tikhonov functionals with sparsity constraints*, Inverse Problems, 24 (2008), p. 035007.

[29] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, Pitman Publishing Limited, 1985.

[30] T. Hein, *Tikhonov regularization in Banach spaces — improved convergence rates results*, Inverse Problems, 25 (2009), p. 035002.

[31] B. Hofmann, B. Kaltenbacher, C. Poeschl, and O. Scherzer, *A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators*, Inverse Problems, 23 (2007), pp. 987–1010.

[32] L. Hörmander, *The Analysis of Linear Partial Differential Operators I*, Springer-Verlag, 1990.

[33] H. Lee, A. Battle, R. Raina, and A. Y. Ng, *Efficient sparse coding algorithms*, in Advances in Neural Information Processing Systems 19, B. Schölkopf, J. Platt, and T. Hoffman, eds., MIT Press, Cambridge, MA, 2007, pp. 801–808.

[34] J. Lindenstrauss and L. Tzafriri, *Classical Banach Spaces II. Function Spaces*, Springer, 1979.

[35] D. A. Lorenz, *Convergence rates and source conditions for Tikhonov regularization with sparsity constraints*, Journal of Inverse and Ill-Posed Problems, 16 (2008), pp. 463–478.

[36] D. A. Lorenz and D. Trede, *Optimal convergence rates for Tikhonov regularization in Besov scales*, Inverse Problems, 24 (2008), p. 055010.

[37] ———, *Greedy deconvolution of point-like objects*, in Proceedings of SPARS'09, April 2009.

[38] Y. Mao, B. Dong, and S. Osher, *A nonlinear PDE-based method for sparse deconvolution*, Multiscale Modeling and Simulation, 8 (2010), pp. 965–976.

[39] L. M. Mugnier, T. Fusco, and J.-M. Conan, *MISTRAL: a myopic edge-preserving image restoration method, with application to astronomical adaptive-optics-corrected long-exposure images*, J. Opt. Soc. Am. A, 21 (2004), pp. 1841–1854.

[40] Y. E. Nesterov, *A method of solving a convex programming problem with convergence rate $O(1/k^2)$*, Soviet Math. Dokl., 27 (1983), pp. 372–376.

[41] A. Neubauer, *On enhanced convergence rates for Tikhonov regularization of nonlinear ill-posed problems in Banach spaces*, Inverse Problems, 25 (2009), p. 065009.

[42] E. Resmerita and O. Scherzer, *Error estimates for non-quadratic regularization and the relation to enhancement*, Inverse Problems, 22 (2006), pp. 801–814.

[43] O. Scherzer and B. Walch, *Sparsity regularization for Radon measures*, in Scale Space and Variational Methods in Computer Vision, X.-C. Tai, K. Morken, M. Lysaker, and K.-A. Lie, eds., vol. 5567 of Lecture Notes in Computer Science, Springer-Verlag, 2009, pp. 452–463. Proceedings of the Second International Conference, SSVM 2009.

[44] G. Stadler, *Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices*, Computational Optimization and Applications, 44 (2009), pp. 159–181.
[45] A. S. Stern, D. L. Donoho, and J. C. Hoch, *NMR data processing using iterative thresholding and minimum l^1-norm reconstruction*, Journal of Magnetic Resonance, 188 (2007), pp. 295–300.
[46] A. N. Tikhonov, A. S. Leonov, and A. G. Yagola, *Nonlinear ill-posed problems*, vol. 1, Chapman & Hall, 1998.
[47] Z. B. Xu and G. F. Roach, *Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces*, J. Math. Anal. Appl., 157 (1991), pp. 189–210.
[48] C. Zălinescu, *Convex analysis in general vector spaces*, World Scientific, 2002.