ESAIM: COCV 19 (2013) 190–218
DOI: 10.1051/cocv/2011205

ESAIM: Control, Optimisation and Calculus of Variations
www.esaim-cocv.org

INVERSE PROBLEMS IN SPACES OF MEASURES

Kristian Bredies[1] and Hanna Katriina Pikkarainen[2]

Abstract. The ill-posed problem of solving linear equations in the space of vector-valued finite Radon measures with Hilbert space data is considered. Approximate solutions are obtained by minimizing the Tikhonov functional with a total variation penalty. The well-posedness of this regularization method and further regularization properties are mentioned. Furthermore, a flexible numerical minimization algorithm is proposed which converges subsequentially in the weak* sense and with rate $\mathcal{O}(n^{-1})$ 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.

Mathematics Subject Classification. 65J20, 46E27, 49M05.

Received November 16, 2010. Revised October 17, 2011.
Published online March 27, 2012.

1. Introduction

This paper is concerned with solving ill-posed linear inverse problems for measure-valued unknowns. While the basic regularization theory of inverse problems is well-established in Hilbert spaces [22] and for linear problems in Banach spaces [34], recent interest focuses on specific linear and non-linear problems usually set in Banach spaces [4,9,21,31,32,38,43,44,49]. One particular research topic is regarding the recovery of sparse unknowns, *i.e.*, exploiting the knowledge that the sought object is, in some sense, composed of only a few elements. This is most commonly formulated in the discrete setting imposing an ℓ^0 constraint [10,15,16,26]. Relaxing this constraint 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 [3, 5–7, 12, 14, 19, 23, 28, 29, 35, 37]. 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 [14, 17, 45, 48]. The present work aims at establishing a framework for this continuous setting in terms of studying general spaces of Radon measures both as the solution space for linear inverse problems as well as the underlying space for numerical algorithms.

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

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

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

Article published by EDP Sciences

© EDP Sciences, SMAI 2012

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

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

$$K^*\mu = f^0 \tag{1.1}$$

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 adjoints 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 linear combinations of delta peaks, which can be considered as an analogue of sparsity with respect to the possibly continuous index set Ω.

Solving the inverse problem (1.1) 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 nearby 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

$$\min_{\mu \in \mathcal{M}(\Omega, \mathbb{R}^m)} \frac{\|K^*\mu - f^\delta\|^2}{2} + \alpha \|\mu\|_{\mathcal{M}} \tag{1.2}$$

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 the 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 on open subsets of $\mathbb{R}^d$ has been recently introduced in [45] 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 on separable, locally compact metric spaces, is dimension-independent and includes the inverse problems considered in [45] as a special case. 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 but we will obtain these results as a special case; namely with K^* in (1.2) being the solution operator of an elliptic equation. Finally, we mention that in [39], 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 [41] or when mass spectrograms are acquired in which a substance contains multiple isotopes [8]. Moreover, the setting might be applied to formulate and solve the problem of finding an optimal placement for control devices or point actuators for partial differential equations [46].

The outline of the paper is as follows. In Section 2 a short introduction to the space of finite Radon measures is given. Section 3 introduces the functional-analytic framework for solutions of (1.2). 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 4, the minimization problem (1.2) is viewed as a regularization method for the inverse problem (1.1). Convergence of the respective solutions of (1.2) as $\delta \to 0$ is studied. 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 is derived. 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 5 proposes a practical numerical algorithm which produces a sequence of approximations

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

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 6. 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 7, summarizing the results.

2. Preliminaries

Before going into the precise definition of the inverse problem (1.1), we introduce some notation and recall some facts from measure theory.

We are dealing with vectors in $\mathbb{R}^m$ for $m \geq 1$. We equip $\mathbb{R}^m$ with a norm $|\cdot|$ and the associated dual norm $|\cdot|_*$ with respect to the dual pairing induced by the scalar product $\sigma \cdot \tau = \sum_{i=1}^m \sigma_i \tau_i$. Furthermore, we denote by $|\cdot|_2$ the Euclidean norm in $\mathbb{R}^m$.

Recall the following generalization of the sign mapping to $\mathbb{R}^m$: 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 $\operatorname{Sgn}(\tau) = \{\tau/|\tau|_2\}$ for $\tau \neq 0$ and $\operatorname{Sgn}(0) = \{|\sigma|_2 \leq 1\}$. In particular, in dimension one it holds that $\operatorname{Sgn}(s) = 1$ for $s > 0$, $\operatorname{Sgn}(s) = -1$ for $s < 0$ and $\operatorname{Sgn}(0) = [-1, 1]$. Also, for $|\tau| = \sum_{i=1}^m |\tau_i|$ it turns out that $\sigma \in \operatorname{Sgn}(\tau)$ if and only if $\sigma_i \in \operatorname{Sgn}(\tau_i)$ for all $i = 1, \ldots, m$.

In this paper, Ω is always a separable locally compact metric space with a metric ρ. In some cases, we assume that Ω is a subset of $\mathbb{R}^d$, $d \geq 1$. Then the metric in $\Omega \subset \mathbb{R}^d$ is induced by the Euclidean norm $|\cdot|_2$ if not mentioned otherwise. In $\mathbb{R}^d$, the boundary of a set E is denoted by ∂E, χ_E is the characteristic function of E, *i.e.*, $\chi_E(x) = 1$ if $x \in E$ and $\chi_E(x) = 0$ if $x \notin E$, and $\operatorname{dist}(E, F)$ is the distance between sets E and F defined by

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

For the measure-theoretic part, we follow the notation of [2]. A positive measure ν is a σ-additive set function defined on the Borel sets of Ω (with respect to the topology induced by the metric ρ) with values in $[0, \infty]$ and $\nu(\emptyset) = 0$. We denote by $L^p_\nu(\Omega, \mathbb{R}^m)$, $1 \leq p < \infty$, the Banach space of all measurable functions $g : \Omega \to \mathbb{R}^m$ such that its p-norm is finite, *i.e.*,

$$\|g\|_p = \left(\int_\Omega |g(\omega)|^p \ \mathrm{d}\nu(\omega) \right)^{1/p} < \infty.$$

Two functions in $L^p_\nu(\Omega, \mathbb{R}^m)$ are identified if they are equal ν-almost everywhere. Let $\mathcal{L}^d$ be the Lebesgue measure in $\mathbb{R}^d$. For $\Omega \subset \mathbb{R}^d$, we use the notation $L^p_{\mathcal{L}^d}(\Omega, \mathbb{R}^m) = L^p(\Omega, \mathbb{R}^m)$ (and $L^p(\Omega)$ if $m = 1$) for all $1 \leq p < \infty$. The space $L^2(\Omega, \mathbb{R}^m)$ is a Hilbert space with the inner product

$$\langle u, \, v \rangle_2 = \sum_{i=1}^m \int_\Omega u_i(x) v_i(x) \, \mathrm{d}x$$

and the corresponding induced norm $\|\cdot\|_2$.

A *finite Radon measure* μ is a σ-additive $\mathbb{R}^m$-valued set function defined on the Borel sets of Ω with $\mu(\emptyset) = 0$. The *total-variation measure* of a finite Radon measure μ is the positive measure defined by

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

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

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 the total variation norm

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

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 the space $\mathcal{M}(\Omega, \mathbb{R}^m)$, which is known as the Riesz theorem, is its characterization as a dual space. To state this theorem, introduce the space

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

which is the completion of continuous compactly supported functions with respect to the supremum norm. For $m = 1$ we denote it by $\mathcal{C}_0(\Omega)$. The Riesz theorem now states that the dual space $(\mathcal{C}_0(\Omega, \mathbb{R}^m))^* = \mathcal{M}(\Omega, \mathbb{R}^m)$ which also implies the completeness of the space of finite Radon measures. The corresponding dual paring is defined by

$$\langle \mu, u \rangle = \sum_{i=1}^m \int_\Omega u_i \, \mathrm{d}\mu_i$$

for each $\mu \in \mathcal{M}(\Omega, \mathbb{R}^m)$ and $u \in \mathcal{C}_0(\Omega, \mathbb{R}^m)$.

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 the mapping $\mu \mapsto \|\mu\|_{\mathcal{M}}$ is lower semi-continuous in this weak* sense. Moreover, every bounded sequence in $\mathcal{M}(\Omega, \mathbb{R}^m)$ has a weak* convergent subsequence. We will subsequently deal with δ-peaks (or Dirac measures) which are defined for $x \in \Omega$ and $m = 1$ by: $\langle \delta_x, u \rangle = u(x)$ for $u \in \mathcal{C}^0(\Omega)$.

Remark 2.1. Let Ω be an open non-empty subset of $\mathbb{R}^d$. In this case, there are some weakly* dense subsets of $\mathcal{M}(\Omega, \mathbb{R}^m)$ which are of interest later on:

(i) the space $L^2(\Omega, \mathbb{R}^m)$ with the injection $w \mapsto w\mathcal{L}^d$;
(ii) the space of all finite linear combinations of δ-peaks:

$$\left\{ \mu = \sum_{i=1}^N u_i \delta_{x_i} \ : \ N \in \mathbb{N}, \ u_i \in \mathbb{R}^m, \ x_i \in \Omega, \ i = 1, \dots, N \right\}.$$

The proof of these density statements can be found in the appendix.

Finally, introduce the notation $\mathcal{L}(X, Y)$ for the set of linear and continuous mappings $K : X \to Y$. We denote by $\|K\| = \sup \{\|Kx\|_Y \ : \ \|x\|_X \leq 1\}$ the corresponding *operator norm*.

3. Minimization in spaces of measures

This section is devoted to the analysis of the minimization problem (1.2). We start with defining more precisely the setting for the inverse problem (1.1). Let H be a Hilbert space with the inner product $\langle \cdot, \cdot \rangle_H$ and the induced norm $\|\cdot\|$. We consider the inversion of K^* which is the adjoint of a continuous linear 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. Throughout this section,

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

we denote by f some given data in H. It can be either the exact data f^0 or the noisy data f^δ. For $\alpha > 0$, we use the notation T_α for the Tikhonov functional

$$T_\alpha(\mu) = \frac{\|K^*\mu - f\|^2}{2} + \alpha\|\mu\|_{\mathcal{M}}. \tag{3.1}$$

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

Proposition 3.1. *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*

$$\min_{\mu \in \mathcal{M}(\Omega, \mathbb{R}^m)} \frac{\|K^*\mu - f\|^2}{2} + \alpha\|\mu\|_{\mathcal{M}} \tag{3.2}$$

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

$$\langle K^*\mu^n, v\rangle_H = \sum_{i=1}^m \int_\Omega Kv \, \mathrm{d}\mu_i^n \to \sum_{i=1}^m \int_\Omega Kv \, \mathrm{d}\mu_i = \langle K^*\mu, v\rangle_H \tag{3.3}$$

for all $v \in H$. Denote by $T_\alpha(\mu) = S(\mu) + \alpha R(\mu)$ with

$$S(\mu) = \tfrac{1}{2}\|K^*\mu - f\|^2 \qquad \text{and} \qquad R(\mu) = \|\mu\|_{\mathcal{M}}. \tag{3.4}$$

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. □

Remark 3.2. As was seen in the above proof, the adjoint $K^* : \mathcal{M}(\Omega, \mathbb{R}^m) \to H$ is weak*-to-weak (sequentially) continuous (*cf.* (3.3)). This property was important for showing the weak* lower semi-continuity of associated Tikhonov functionals.

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

$$\begin{aligned}\lim_{n\to\infty} \langle A^*u, v^n\rangle_{Y^{**}\times Y^*} &= \lim_{n\to\infty} \langle u, Av^n\rangle_{X^{**}\times X^*} \\ &= \langle u, Av\rangle_{X^{**}\times X^*} = \langle A^*u, v\rangle_{Y^{**}\times Y^*}\end{aligned}$$

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

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

Since in the studied case $X = H$ is a Hilbert space, the weak*-to-weak* sequential continuity of A is equal to the weak*-to-weak sequential continuity, therefore, these mappings coincide with the adjoints mappings.

Before proceeding to the analysis of minimizers of (3.2), 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 3.3. Let Ω be a bounded domain in $\mathbb{R}^d$ 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 continuous functions. First, we see that the translation of a L^2-function, *i.e.*, $x \mapsto k(\cdot - x)$, is a continuous operation mapping $\Omega'' \to L^2(\mathbb{R}^d)$ (see, for instance, [1], Thm. 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 $\mathrm{supp}\, K_\varepsilon v^\varepsilon \subset \overline{\Omega_\varepsilon - \Omega'_\varepsilon} \subset\subset \Omega''$. 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 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,

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

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 3.1 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 $\mu \in \mathcal{M}(\Omega'')$

$$\langle K^*\mu,\, v\rangle_2 = \int_{\Omega''} \int_\Omega k(y - x)v(y)\ \mathrm{d}y\ \mathrm{d}\mu(x) = \int_\Omega \int_{\Omega''} k(y - x)\ \mathrm{d}\mu(x)v(y)\ \mathrm{d}y$$

by Fubini's theorem, so $K^*\mu = \mu * k$ in the sense of the convolution of measures. It turns out that (3.2) 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 3.4. 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

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

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

$$\begin{aligned} -\operatorname{div}(A\nabla u) &= v \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{3.5}$$

From the regularity theory of elliptic equations (see, for instance, [30], Chap. 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 3.1 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

$$\begin{aligned} -\operatorname{div}(A\nabla f) &= \mu \quad \text{in } \Omega, \\ f &= 0 \quad \text{on } \partial\Omega \end{aligned} \tag{3.6}$$

in the weak sense of Stampacchia [47]: Indeed, for each $v \in \mathcal{C}_0(\Omega)$ we have $u = Kv \in H_0^1(\Omega) \cap \mathcal{C}_0(\Omega)$. Moreover, denoting L^* as the formal adjoint of the differential operator $L = -\operatorname{div}(A\nabla)$ it follows that $L^*u = v$ in the weak sense. Hence, for $\mu \in \mathcal{M}(\Omega)$ and $f = K^*\mu$ it holds that:

$$\int_\Omega L^*uf \,\mathrm{d}x = \int_\Omega vf \,\mathrm{d}x = \langle K^*\mu,\, v\rangle_2 = \langle \mu,\, Kv\rangle = \int_\Omega u \,\mathrm{d}\mu,$$

implying that f satisfies (3.6) in the sense of Definition 9.1 in [47]. Moreover, as Ω satisfies the uniform cone property, it is an $H_0^1(\Omega)$-admissible domain according to Definition 6.2 in [47]. It is immediate that a weak solution of (3.6) has to be unique, hence Theorem 9.1 in [47] implies that actually $f \in W_0^{1,q}(\Omega)$ for each $1 \leq q < d/(d-1)$. Equation (3.6) is therefore also satisfied in the distributional sense in Ω with homogeneous Dirichlet boundary values attained in the sense of traces.

Next, we will derive optimality conditions for minimizers of (3.2) with the help of convex duality (see, for instance, [20, 52]). The usual setting, 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 (3.2) 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. Here, the indicator functional I of a set E is defined by $I_E(x) = 0$ if $x \in E$ and $I_E(x) = \infty$ if $x \notin E$. Also, the subdifferential and the Fenchel conjugate of a functional F are denoted by ∂F and F^*, respectively.

Proposition 3.5. *Let $K : H \to \mathcal{C}_0(\Omega, \mathbb{R}^m)$ be linear and continuous, $f \in H$ and $\alpha > 0$. The problem*

$$\min_{v \in H} \frac{\|v - f\|^2}{2} + I_{\{v \in H \,:\, \|Kv\|_\infty \leq \alpha\}}(v) \tag{3.7}$$

is equivalent to a Fenchel predual problem of (3.2), i.e., the Fenchel dual problem of (3.7) exists and possesses the same solutions as (3.2).

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 $R_0(u) = I_{\{\|u\|_\infty \leq \alpha\}}(u)$ reads as

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

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

where R is defined in (3.4). Now verify that all prerequisites for the validity of the Fenchel duality formula according to [51], 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 (3.2). □

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

Proposition 3.6. *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 (3.2) if and only if $w^* = -K(K^*\mu^* - f)$ satisfies $\|w^*\|_\infty \leq \alpha$ and each polar decomposition $\mu^* = \sigma^*|\mu^*|$ obeys*

$$\sigma^* \cdot w^* = \alpha \quad |\mu^*|\text{-almost everywhere.} \tag{3.8}$$

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

Proof. From Fenchel duality and Proposition 3.5 one knows that $(v^*, \mu^*) \in H \times \mathcal{M}(\Omega, \mathbb{R}^m)$ are solutions of (3.7) and (3.2), 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 3.5. 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

$$\int_\Omega (w - w^*) \cdot \sigma^* \, \mathrm{d}|\mu^*| \leq 0 \quad \text{for all } w \in \mathcal{C}_0(\Omega, \mathbb{R}^m),\ \|w\|_\infty \leq \alpha. \tag{3.9}$$

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 (3.9), 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 3.7. If the vector norm $|\cdot|$ is strictly convex, *i.e.*, $|u|, |v| \leq 1$ implies $|u + v| < 2$ for $u \neq v$, 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, $\sigma^*(x) = J(w^*(x))$ for $|\mu^*|$-almost all $x \in \Omega$ where $J : \mathbb{R}^m \to \mathbb{R}^m$ is the $|\cdot|_*$-$|\cdot|$-duality mapping

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

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 (3.8),

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

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

On the other hand, each Borel measurable 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: $\sigma^*(x) = J\left(w^*(x)\right)$ 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 3.8. 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 (3.2) 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

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

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 (3.8) is an equivalent description of the above in terms of measures.

Remark 3.9. The last remark shows that the situation (3.2) 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 Tikhonov functional with L^1-regularization (with some technical effort). As we have seen in Proposition 3.1, considering $\mathcal{M}(\Omega, \mathbb{R}^m)$ instead of $L^1(\Omega, \mathbb{R}^m)$ provides the compactness needed in the existence proof.

4. Regularization properties

In this section we interpret the minimization problem (1.2) as a Tikhonov regularization method for the inverse problem (1.1). Even though our problem is covered by the existing theory of regularization in Banach spaces [9, 32], we want to discuss this matter shortly for the sake of consistence. We point out that the minimization problem (1.2) is well-posed and approximates the inverse problem (1.1) in the sense usually considered in regularization theory.

Like in the previous section, we denote by f^0 the exact data and by f^δ noisy data. We assume that there exists a solution of (1.1), *i.e.*, f^0 belongs to the range of the operator K^*. If the null space of K^* is not trivial,

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

the inverse problem (1.1) may have multiple solutions. Our interest is in the following special solutions (if they exist):

Definition 4.1. An element $\mu^\dagger \in \mathcal{M}(\Omega, \mathbb{R}^m)$ is called a *minimum norm solution* of (1.1) if

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

We call a minimizer of the Tikhonov functional T_α with data f^δ a regularized solution of (1.1) (see (3.1) for the definition). 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 regularized solution is a good approximation of a solution of the inverse problem (1.1) if there exists a choice of the regularization parameter α such that a regularized solution converges to a minimum norm solution as the noise level tends to zero.

In the setup of the paper, 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, see [2], Theorem 1.59;
- every bounded sequence in $\mathcal{M}(\Omega, \mathbb{R}^m)$ has a weak* convergent subsequence, see also [2], Theorem 1.59.

Accordingly, the inverse problem (1.1) is a specific example of problems studied in [32]. Therefore the following results are valid:

- a minimum norm solution $\mu^\dagger$ of (1.1) exists, see [32], Theorem 3.4;
- *existence*: the Tikhonov functional T_α admits a minimizer (Prop. 3.1);
- *uniqueness*: the minimizer of T_α is unique if K^* is injective (Prop. 3.1);
- *stability*: for a norm-convergent sequence of the data every minimizing sequence of the corresponding Tikhonov functionals has a subsequence which converges in the weak* sense. The limit of each weak* convergent subsequence is a minimizer of the Tikhonov functional with the limit data, see [32], Theorem 3.2;
- *convergence*: with the parameter choice rule

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

  where $\alpha(\cdot)$ is a monotonically increasing function, regularized solutions converge subsequentially in the weak* sense to a minimum norm solution as $\delta \to 0$. If, in addition, the minimum norm solution $\mu^\dagger$ is unique, the convergence is sequential, not only subsequential, see [32], Theorem 3.5.

The well-posedness of the minimization of the Tikhonov functional is guaranteed by the existence, uniqueness and stability results above. Furthermore, the convergence of regularized solutions to a minimum norm solution is obtained (subsequentially in the weak* sense) with an *a-priori* parameter choice rule. Hence the minimization problem (1.2) is a proper regularization method for the inverse problem (1.1).

For qualitative estimates for the convergence of regularized solutions 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

$$\exists\, h \in H : \quad K^{**}h \in \partial\|\cdot\|_{\mathcal{M}}(\mu^\dagger). \tag{4.1}$$

The subdifferential of the norm $\|\cdot\|_{\mathcal{M}}$ is $\partial\|\cdot\|_{\mathcal{M}}(\mu) = \{\varphi \in (\mathcal{M}(\Omega, \mathbb{R}^m))^* : \langle\varphi, \mu\rangle_{\mathcal{M}^*\times\mathcal{M}} = \|\mu\|_{\mathcal{M}}$ 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

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

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

$$\exists\, h \in H : \quad \langle \mu^\dagger, Kh\rangle = \|\mu^\dagger\|_{\mathcal{M}} \quad \text{and} \quad \|Kh\|_\infty = 1. \tag{4.2}$$

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

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

Definition 4.2. 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

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

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

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

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

- *Convergence rates*: with the parameter choice rule $\alpha \sim \delta$, in the Bregman distance related to the total variation norm between a regularized solution and a minimum norm solution satisfying the source condition (4.2) there exists an element which converges with rate $\mathcal{O}(\delta)$, see [9], Theorem 2.

Remark 4.3. The source condition of the form (4.1) 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 $(\mathcal{M}(\Omega, \mathbb{R}^m))^*$ is complicated and therefore at first glance the source condition is hard to verify. In this paper, the form of the inverse problem (1.1), *i.e.*, the existence of the predual operator K, simplifies the source condition, as has been seen in (4.2). Note that for this particular problem the dual space $(\mathcal{M}(\Omega, \mathbb{R}^m))^*$ or its norm are not needed in the source condition. This may not be the case for all inverse problems involving finite Radon measures.

We can further modify the source condition (4.2). 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 (4.2), $\|Kh\|_\infty = 1$. 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

$$\exists\, h \in H : \quad \begin{cases} (Kh)(x) \cdot \sigma^\dagger(x) = 1 & \text{for } |\mu^\dagger|\text{-almost every } x \in \Omega, \\ |(Kh)(x)| \le 1, & x \in \Omega. \end{cases} \tag{4.3}$$

Remark 4.4. Let us now consider an inverse problem $K^*\mu = f$ where $\mu \in \mathcal{M}(\Omega)$ and $f \in H$, *i.e.*, the case $m = 1$. 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 $\mu^\dagger = \sigma^\dagger|\mu^\dagger|$. Let us redefine the value of $\sigma^\dagger$ to be 1 in the $|\mu^\dagger|$-null set $\{x \in \Omega : |\sigma^\dagger(x)| \neq 1\}$. 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 (4.3) 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

$$\begin{cases} |\mu^\dagger|(\Omega^\dagger_+ \setminus \widetilde{\Omega}_+) = 0, \\ |\mu^\dagger|(\Omega^\dagger_- \setminus \widetilde{\Omega}_-) = 0, \end{cases} \quad \text{and} \quad \begin{cases} (Kh)(x) = 1, & x \in \widetilde{\Omega}_+, \\ (Kh)(x) = -1, & x \in \widetilde{\Omega}_-, \\ |(Kh)(x)| \le 1, & x \in \Omega. \end{cases} \tag{4.4}$$

Since Kh is a $\mathcal{C}_0(\Omega)$-function, in view of (4.4) 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.

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

Example 4.5 (continued)**.** Consider now the elliptic boundary value problem (3.6). Let the assumptions made in Example 3.4 be valid. Let $f \in L^2(\Omega)$ and $\mu^\dagger \in \mathcal{M}(\Omega)$ be the (unique) solution of the inverse problem $K^*\mu = f$ where K^* is the solution operator for the elliptic problem (3.6).

With the notations of Remark 4.4, let us suppose that there exist sets $\widetilde{\Omega}_+ \subset \Omega_+^\dagger$ and $\widetilde{\Omega}_- \subset \Omega_-^\dagger$ such that

$$\begin{cases} |\mu^\dagger|(\Omega_+^\dagger \setminus \widetilde{\Omega}_+) = 0, \\ |\mu^\dagger|(\Omega_-^\dagger \setminus \widetilde{\Omega}_-) = 0, \end{cases} \qquad \text{and} \qquad \begin{cases} \operatorname{dist}(\widetilde{\Omega}_+, \widetilde{\Omega}_-) > 0, \\ \operatorname{dist}(\widetilde{\Omega}_+, \partial\Omega) > 0, \\ \operatorname{dist}(\widetilde{\Omega}_-, \partial\Omega) > 0. \end{cases} \tag{4.5}$$

According to Remark 4.4, the existence of sets $\widetilde{\Omega}_+$ and $\widetilde{\Omega}_-$ satisfying requirements (4.5) is a necessary condition for $\mu^\dagger$ to fulfill the source condition (4.4).

Using, for instance, Theorem 1.4.1 in [33], we can construct functions g_+, $g_- \in \mathcal{C}^\infty(\Omega)$ with disjoint compact supports such that $0 \leq g_+, g_- \leq 1$,

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

Define $u = g_+ - g_-$ and $v = -\operatorname{div}(A\nabla u)$ in the weak sense. Then $u = Kv$ and u satisfies the conditions in (4.4) 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 (4.5) is actually the sufficient condition for $\mu^\dagger$ to fulfill the source condition (4.4) in the case of the elliptic problem (3.6).

Note that for sparse solutions, *i.e.*,

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

the sets $\widetilde{\Omega}_+$ and $\widetilde{\Omega}_-$ with properties (4.5) always exist and hence, sparse solutions satisfy the source condition (4.4).

5. A numerical minimization algorithm

The following section aims at proposing a numerical algorithm for the iterative minimization of the Tikhonov functional (3.2) 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 [36,50]. On the other hand, Radon measures are well-studied objects and many notions of convergence are available (see, for instance, [24,27]) 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.*, solving for $A \in \mathcal{L}\left(\ell^2, H\right)$ and $f \in H$ the problem

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

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

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

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

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

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

This algorithm is known to converge in the sequence-space setting, in many cases at a q-linear rate [6, 12, 14]. 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 $u \in \ell^2$.

In the existing literature, this defect is often circumvented by going into the discrete setting as follows. Consider again problem (3.2) in the situation of Proposition 3.1. We now assume that the sought measure μ consists of finitely many fixed delta peaks located at $\Omega_h = \{x_1, \dots, x_N\} \subset \Omega$ and coefficients $u_k \in \mathbb{R}^m$:

$$\mu = \sum_{k=1}^{N} u_k \delta_{x_k} = \sum_{k=1}^{N}\sum_{j=1}^{m} u_{k,j}(e_j \delta_{x_k})$$

with e_j being the j-th unit vector in $\mathbb{R}^m$. Therefore, writing

$$u = (u_1, \dots, u_N) \quad , \quad Au = \sum_{k=1}^{N}\sum_{j=1}^{m} u_{k,j} K^*(e_j \delta_{x_k})$$

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 [25]. 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 [45].

Our approach stays, in contrast to the above, in the continuous domain and incorporates a way of adding and removing delta peaks at arbitrary points of Ω. It realizes the above finite-dimensional soft-thresholding step as a part of the algorithm but also provides a mechanism to iteratively modify the set of peak positions Ω_h.

Subsequently, we investigate an algorithm for the minimization of the Tikhonov functional T_α in (3.2). In order to formulate the algorithm, we need some prerequisites. First, define, for a $t \geq 0$,

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

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

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

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

$$\mathcal{S}_\alpha(w) = \operatorname{sgn}(w)\max(0, |w| - \alpha),$$

for instance. Finally, denote by $C_{|\cdot|}$ a constant for which $|w| \leq C_{|\cdot|}|w|_2$ holds for all $w \in \mathbb{R}^m$.

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

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

Algorithm 5.1.

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
$$v^n = \begin{cases} 0 & \text{if } |w^n(x_*^n)|_* \leq \alpha, \\ \alpha^{-1} M_0 \|w^n\|_\infty \operatorname{sgn}_*(w^n(x_*^n)) & \text{if } |w^n(x_*^n)|_* > \alpha. \end{cases}$$
4. Evaluate, with division by zero resulting in ∞,
$$s_n = \min\left\{1, \frac{\alpha \sum_{j=1}^{J_n} |u_j^n| - \varphi(|v^n|) - \langle \mu^n - \nu^n, w^n \rangle}{\|K^*(\mu^n - \nu^n)\|^2}\right\},$$
stop if $s_n = 0$ with μ^n being a solution. Otherwise, define $u_j^{n+1/2} \in \mathbb{R}^m$, $x_j^{n+1} \in \Omega$ for $j = 1, \ldots, J_n + 1$ and $\mu^{n+1/2}$ by
$$\mu^{n+1/2} = \mu^n + s_n(\nu^n - \mu^n) = \sum_{j=1}^{J_n+1} u_j^{n+1/2} \delta_{x_j^{n+1}}.$$
5. Set $s_{n+1/2} = s_* / \left(C_{|\cdot|}^2 (J_n + 1)\right)$ and compute
$$u_j^{n+1} = \mathcal{S}_{\alpha s_{n+1/2}}\left(u_j^n - s_{n+1/2}\left(K(K^*\mu^{n+1/2} - f)\right)(x_j^{n+1})\right)$$
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 (see [5, 7]) and the well-known iterative soft-thresholding method (or, more generally, a generalized gradient projection method [6]). 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. In order to make the generalized conditional gradient method well-defined, we change the term $\alpha\|\cdot\|_\mathcal{M}$ to $\varphi \circ \|\cdot\|_\mathcal{M}$ in the Tikhonov functional. This does, however, not affect the solutions.

Lemma 5.2. *The minimizers of T_α are exactly the minimizers of*
$$\tilde{T}_\alpha(\mu) = \frac{\|K^*\mu - f\|^2}{2} + \varphi(\|\mu\|_\mathcal{M}).$$

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

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

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

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. □

For carrying out the convergence analysis, we introduce some notation which will be common to the following propositions and lemmas:

Definition 5.3.

(i) Let μ^* be an arbitrary but fixed minimizer of T_α;
(ii) denote by $r(\mu) = T_\alpha(\mu) - T_\alpha(\mu^*)$ the *functional distance*;
(iii) introduce the *primal-dual gap*

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

Our convergence analysis regarding steps 2–4 is based on the presentation in [5] which is carried out in Hilbert spaces. As we will see, some of the arguments can be transferred to the Banach space setting we consider here. The first step consists in characterizing the ν^n in step 3 as the solution of a minimization problem.

Proposition 5.4. *For a $w \in \mathcal{C}_0(\Omega, \mathbb{R}^m)$ a minimizer of*

$$\min_{\nu \in \mathcal{M}(\Omega, \mathbb{R}^m)} \; -\langle \nu,\, w\rangle + \varphi(\|\nu\|_{\mathcal{M}}) \tag{5.1}$$

is given by

$$\nu = \begin{cases} 0 & \text{if } \|w\|_\infty \leq \alpha, \\ \alpha^{-1} M_0 \|w\|_\infty \mathrm{sgn}_*\left(w(x)\right) \delta_x & \text{otherwise} \end{cases} \tag{5.2}$$

with $x \in \Omega$ such that $|w(x)|_* = \|w\|_\infty$.

Moreover, the set-valued operator taking w to all solutions ν of (5.1) is bounded on bounded sets.

Proof. Note that since φ is convex on $[0, \infty[$, a minimizer ν of (5.1) 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

$$\langle \nu,\, w\rangle = \|\nu\|_{\mathcal{M}}\|w\|_\infty \quad \text{and} \quad \|w\|_\infty \begin{cases} \leq \alpha & \text{if } \nu = 0, \\ = \alpha & \text{if } 0 < \|\nu\|_{\mathcal{M}} \leq M_0, \\ = \frac{\alpha\|\nu\|_{\mathcal{M}}}{M_0} & \text{else.} \end{cases} \tag{5.3}$$

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, $\nu = \alpha^{-1} M_0 \|w\|_\infty \mathrm{sgn}_*\left(w(x)\right) \delta_x$ satisfies $\|\nu\|_{\mathcal{M}} > M_0$ and the definition of the sign operator yields

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

as well as $\frac{\alpha\|\nu\|_{\mathcal{M}}}{M_0} = \|w\|_\infty$. Consequently, ν is a minimizer of (5.1).

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

Eventually, let $w \in \mathcal{C}_0(\Omega, \mathbb{R}^m)$ and $\nu \in \mathcal{M}(\Omega, \mathbb{R}^m)$ be a minimizer of (5.1). According to (5.3), 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. □

This characterization allows us to obtain a useful estimate on the primal-dual gap for the choice w according to step 2.

Lemma 5.5. *For $\mu \in \mathcal{M}(\Omega, \mathbb{R}^m)$, $w = -K(K^*\mu - f)$ and ν a solution of (5.1), it holds that $G(\mu, \nu) \geq r(\mu)$.*

Proof. 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 (5.1), to

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

Hence the inequality holds as stated. □

Combining the latter two results with what is done in step 4, we can further derive a descent estimate for the functional.

Proposition 5.6. *Let $\mu \in \mathcal{M}(\Omega, \mathbb{R}^m)$ with $T_\alpha(\mu) \leq \|f\|^2/2$, $w = -K(K^*\mu - f)$, ν a solution of (5.1) and*

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

Then, the convex combination $\bar{\mu} = \mu + s(\nu - \mu)$ satisfies

$$r(\bar{\mu}) - r(\mu) \leq -c r(\mu)^2$$

with $c > 0$ independent of μ and ν.

Proof. Observe again that $T_\alpha(\mu) \leq \|f\|^2/2 = T_\alpha(0)$ implies $\|\mu\|_{\mathcal{M}} \leq M_0$ 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

$$\varphi\left(\|\bar{\mu}\|_{\mathcal{M}}\right) - \varphi\left(\|\mu\|_{\mathcal{M}}\right) \leq s\left(\varphi\left(\|\nu\|_{\mathcal{M}}\right) - \varphi\left(\|\mu\|_{\mathcal{M}}\right)\right).$$

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

$$\begin{aligned} \tilde{T}_\alpha(\bar{\mu}) - \tilde{T}_\alpha(\mu) &\leq s\left(\langle \mu - \nu, w \rangle + \varphi\left(\|\nu\|_{\mathcal{M}}\right) - \alpha \|\mu\|_{\mathcal{M}}\right) + \frac{s^2 \|K^*(\nu - \mu)\|^2}{2} \\ &= -sG(\mu, \nu) + \frac{s^2 \|K^*(\nu - \mu)\|^2}{2}. \end{aligned} \tag{5.4}$$

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

We now distinguish two cases: first, suppose $\|K^*(\nu-\mu)\|^2 \leq G(\mu,\nu)$ leading to $s=1$ and, consequently, with the help of Lemma 5.5, to

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

With

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

such that $C_1 > 0$, the above becomes

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

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 (5.4), we arrive at

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

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 Prop. 5.4), there follows $\|K^*(\nu-\mu)\|^2 \leq \|K^*\|^2(C_2+M_0)^2 \leq C_3$ for some $C_3 > 0$. Thus

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

In both of the two cases, $\tilde{T}_\alpha(\bar{\mu}) \leq \tilde{T}_\alpha(\mu) = T_\alpha(\mu) \leq T_\alpha(0)$ and hence, $\|\bar{\mu}\|_{\mathcal{M}} \leq M_0$ 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 $r(\bar{\mu}) - r(\mu) \leq -cr(\mu)^2$ is obtained. □

Taking a look at Algorithm 5.1, we see that ν chosen in steps 2–3 corresponds to a minimizing element for (5.1) (see Prop. 5.4), so step 4 produces $\mu^{n+1/2} = \bar{\mu}^n$ according to Proposition 5.6. 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 which is inspired by the presentation in [6].

Proposition 5.7. *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*

$$v_j = \mathcal{S}_{s\alpha}\left(u_j - s\left(B^*K(K^*Bu - f)\right)_j\right), \quad j = 1, \ldots, J, \tag{5.5}$$

satisfies

$$\frac{\|K^*Bv - f\|^2}{2} + \alpha \sum_{j=1}^J |v_j| \leq \frac{\|K^*Bu - f\|^2}{2} + \alpha \sum_{j=1}^J |u_j|. \tag{5.6}$$

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

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

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

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

It is well-known [14, 25] that the soft-thresholding operator $\eta = \mathcal{S}_\beta(\xi)$ is the solution operator to

$$\min_{\eta \in \mathbb{R}^m} \ \frac{|\eta - \xi|_2^2}{2} + \beta |\eta|. \tag{5.7}$$

We nevertheless give a short argumentation for the sake of completeness. It is easily seen, mostly analogous to Proposition 3.5, 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 $w_j = (-B^* K (K^* B u - f))_j$. Since $v_j = \mathcal{S}_{s\alpha}(u_j + s w_j)$ is the solution of the problem (5.7), we can express it with subgradients and use subgradient calculus:

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

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

$$(u_j + s w_j - v_j) \cdot (\eta_j - v_j) \leq s \left(\alpha |\eta_j| - \alpha |v_j| \right).$$

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

$$|u_j - v_j|_2^2 \leq s \left(\alpha |u_j| - \alpha |v_j| + w_j \cdot (u_j - v_j) \right) = s G_j. \tag{5.8}$$

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

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

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

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 $\|B\|$, (5.8), the choice of s and (5.8) again, we get

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

The desired statement now immediately follows. □

By examining the notation, it is easy to see that step 5 of Algorithm 5.1 corresponds to the thresholding step (5.5) and that (5.6) is just $r(\mu^{n+1}) \leq r(\mu^{n+1/2})$. 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 5.1 to minimizers of the Tikhonov functional (3.2).

Theorem 5.8. *Algorithm 5.1 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 5.6 can be applied (note again, that $\mu^{n+1/2}$ corresponds to $\bar{\mu}^n$ by virtue of Prop. 5.4). This leads to $r(\mu^{n+1/2}) \leq r(\mu^n)$ and by virtue of Proposition 5.7, 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 5.6 and 5.7 also yield a $c > 0$ such that

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

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

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

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 Prop. 3.1). In case of uniqueness of μ^*, the statement follows from the usual subsequence argument. □

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

5.1. Acceleration strategies

It is obvious from the proof that Algorithm 5.1 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 5.7, 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 [42] 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 [40], here we adapt it to our general framework.

Proposition 5.9. *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. Let $\mu = \sum_{j=1}^J u_j \delta_{x_j^0}$ with $u_j \neq 0$ and $x_j^0 \in \Omega$ pairwise disjoint. Then, there exists a $t_0 > 0$ for which the system of ordinary differential equations*

$$\begin{cases} \dfrac{\partial x_j(t)}{\partial t} = -\nabla \left(K \left(K^* \sum_{j=1}^J u_j \delta_{x_j(t)} - f \right) \right) (x_j(t))^{\mathrm{T}} u_j \\ x_j(0) = x_j^0, \qquad j = 1, \dots, J \end{cases} \tag{5.9}$$

admits a solution for $t \in]0, t_0[$. Each solution on $]0, t_0[$ satisfies

$$\frac{\partial}{\partial t} \left(T_\alpha \left(\sum_{j=1}^J u_j \delta_{x_j(t)} \right) \right) \leq 0 \qquad \text{for all} \qquad t \in]0, t_0[.$$

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

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

is continuously differentiable with gradient

$$\left(\nabla \Phi(x) \right)_j = \nabla \left(K \left(K^* \mu(x) - f \right) \right) (x_j)^{\mathrm{T}} u_j.$$

For that purpose, let $h = (h_1, \dots, h_J) \in (\mathbb{R}^d)^J$ 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, since the regularization penalty cancels, as

$$\begin{aligned} \frac{1}{\tau} \left(T_\alpha \left(\mu(x + \tau h) \right) - T_\alpha \left(\mu(x) \right) \right) &= \frac{1}{\tau} \left\langle K^* \left(\mu(x + \tau h) - \mu(x) \right), \frac{1}{2} K^* \left(\mu(x + \tau h) + \mu(x) \right) - f \right\rangle_{(\mathbb{R}^d)^J} \\ &= \left\langle \frac{1}{\tau} \left(\mu(x + \tau h) - \mu(x) \right), K \left(K^* \mu(x) - f \right) \right\rangle_{(\mathbb{R}^d)^J} \\ &\quad - \frac{\| K^* \left(\mu(x + \tau h) - \mu(x) \right) \|^2}{2\tau}. \end{aligned}$$

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

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

$$\lim_{\tau\to 0} \frac{1}{\tau} \left\langle \mu(x+\tau h) - \mu(x),\, K\left(K^*\mu(x) - f\right)\right\rangle = \lim_{\tau\to 0} -\sum_{j=1}^{J} u_j \cdot \frac{w(x_j+\tau h_j) - w(x_j)}{\tau} = \sum_{j=1}^{J} -u_j \cdot ((\nabla w)(x_j)h_j)\,.$$

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 $v \in \mathcal{C}_0^1(\Omega, \mathbb{R}^m)$:

$$\begin{aligned}\langle \mu(x+\tau h) - \mu(x),\, v\rangle &= \sum_{j=1}^{J} u_j \cdot (v(x_j+\tau h_j) - v(x_j)) \\ &\leq |\tau| \sum_{j=1}^{J} \int_0^1 |u_j||\nabla v(x_j + s\tau h_j)||h_j|_2 \,\mathrm{d}s \leq C|\tau|\|\nabla v\|_\infty,\end{aligned}$$

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

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

This shows that $x \mapsto T_\alpha(\mu(x))$ 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 (5.9) 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 for each $t \in]0, t_0[$:

$$\frac{\partial(\Phi \circ x)}{\partial t}(t) = \left\langle \frac{\partial x}{\partial t}(t),\, \nabla\Phi\left(x(t)\right)\right\rangle_{(\mathbb{R}^d)^J} = -\sum_{j=1}^{J} |\nabla\Phi\left(x(t)\right)_j|_2^2 \leq 0$$

which completes the proof. □

Consequently, one gets functional descent if one inserts, in step 6, before returning to step 2, the solution of (5.9) 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

$$\begin{aligned}\mu^n &= \sum_{j=1}^{J} u_j \delta_{x_j}, \quad 1 \leq j_0 < j_1 \leq J, \\ \mu^{n+1} &= \begin{cases} (u_{j_0} + u_{j_1})\delta_{\frac{1}{2}(x_{j_0} + x_{j_1})} + \displaystyle\sum_{j\neq j_0,\, j\neq j_1} u_j \delta_{x_j} & \text{if } T_\alpha(\mu^{n+1}) \leq T_\alpha(\mu^n) \\ \mu^n & \text{else} \end{cases}\end{aligned} \tag{5.10}$$

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 5.1 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 (5.9) for some positive time, if possible;
- trying to merge some peaks according to the trial-and-error method (5.10), if possible.

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

6. Numerical computations

To show the performance of Algorithm 5.1, we apply it for a one-dimensional version of the deconvolution problem presented in Example 3.3. 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.*,

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

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

$$(Kv)(x) = \sum_{i=0}^{N-1} v_i k(x - ih_{\mathrm{c}}) \tag{6.1}$$

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 3.1 and in particular of Section 5. It is moreover easy to see that the adjoint corresponds to

$$(K^*\mu)_i = (k * \mu)(ih_{\mathrm{c}}), \qquad i = 0, \ldots, N-1,$$

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

$$T_\alpha(\mu) = \frac{1}{2}\left(\sum_{i=0}^{N-1} ((k * \mu)(ih_{\mathrm{c}}) - f_i)^2\right) + \alpha\|\mu\|_{\mathcal{M}}.$$

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

Algorithm 5.1 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 (6.1)):

$$Kv = \sum_{i=0}^{N-1} v_i \sum_{j=-L+2}^{L-2} k_j \psi_j(\cdot - imh_{\mathrm{f}}) = \sum_{i=0}^{N-1} \sum_{j=-L+2}^{L-2} v_i k_j \psi_{j+im}.$$

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

TABLE 1. Top: the values and positions for $\mu^\dagger$, μ and the corresponding extremal points of w for the first deconvolution experiment found by Algorithm 5.1. Bottom: the outcome of the iterative soft-thresholding algorithm [14] for sparse spike deconvolution as proposed in [17].

Successive peak insertion and thresholding					
$\mu^\dagger = \sum_i u_i^\dagger \delta_{x_i^\dagger}$		$\mu = \sum_i u_i \delta_{x_i}$		$w = -K(K^*\mu - f)$	
$x_i^\dagger$	$u_i^\dagger$	x_i	u_i	x_{ex}	$w(x_{\mathrm{ex}})$
0.100000	1.000000	0.104256	0.401399	0.104256	3.000000
0.500000	2.000000	0.496413	1.262217	0.496412	3.000000
0.600000	5.000000	0.597352	4.541020	0.597352	3.000000
0.800000	1.000000	0.786699	0.698377	0.786699	3.000000

Iterative soft-thresholding for sparse spike deconvolution					
$\mu^\dagger = \sum_i u_i^\dagger \delta_{x_i^\dagger}$		$\mu = \sum_i u_i \delta_{x_i}$		$w = -K(K^*\mu - f)$	
$x_i^\dagger$	$u_i^\dagger$	x_i	u_i	x_{ex}	$w(x_{\mathrm{ex}})$
0.100000	1.000000	0.102957	0.404129	0.105181	3.007119
0.500000	2.000000	0.496774	1.258682	0.494902	3.021706
0.600000	5.000000	0.597460	4.546153	0.597062	2.941770
0.800000	1.000000	0.788443	0.688498	0.785062	3.013024

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 5.1 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. Note that the above arguments heavily depend on the fact that the convolution kernel is a cubic spline. An adaptation to other forward operators K^*, however, seems to be possible if one knows the image of each δ-peak and has a way to compute the extremal points of the dual variable w accurately.

We implemented and tested Algorithm 5.1 for the deconvolution problem. Taking advantage of the special structure of the convolution operator, we also added the ODE-solving step (5.9) as well as the peak-merging step (5.10) to improve the performance. Two experiments were performed. For the first, we chose $N = 100$ data points and $h_{\mathrm{c}} = h_{\mathrm{f}} = 0.01$, a true solution with four delta peaks and a spline convolution kernel approximating a Gaussian with variance $\sigma^2 = 0.05$. To challenge the algorithm and model, the data has been contaminated by quite heavy Gaussian noise (standard variance $\sigma^2 = \frac{1}{4}$). The regularization parameter was chosen $\alpha = 3$ and the iteration was stopped when a decrease in the primal-dual gap $G(\mu^n, \nu^n)$ to the fraction 10^{-9} with respect to the zero initial guess has been observed (which happened after 2647 iterations). As a "sparsifying" post-processing step, we merged the peaks belonging to an extremal point of w by computing the average of the positions weighted by the magnitude. Figure 1 shows, among the used input data, the obtained approximate solution μ, its image $K^*\mu$ and the associated dual variable w. Moreover, peak positions and heights for μ as well as the positions and values of the corresponding extremal points of w are shown in Table 1. We also include, for comparison, the values for the sparse spike deconvolution method proposed in [17] performed with the same data, stopping criterion (which was reached after 18 314 iterations) and postprocessing steps, but with a fixed uniform grid of size $h_{\mathrm{f}} = 0.01$.

One sees that for Algorithm 5.1 with peak moving and merging, the number of peaks is recovered exactly and the positions match quite well. Due to the regularization, however, the recovered peak heights are generally lower than for the true data. This is due to the model: as one can see by inspecting the extremal points of

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

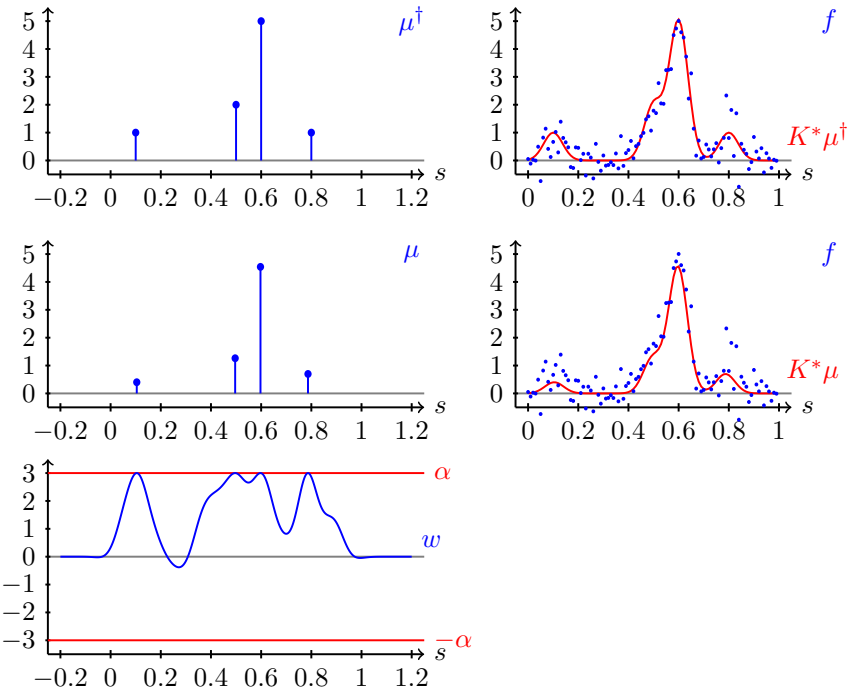


FIGURE 1. First numerical deconvolution experiment. Upper row: the exact solution $\mu^\dagger$ (left), its image $K^*\mu^\dagger$ under convolution and given noisy data f (right). Second row: the numerical approximation μ (left) and its image $K^*\mu$ compared to the data (right). Third row: the corresponding dual variable $w = -K(K^*\mu - f)$.

$w = -K(K^*\mu - f)$, the positions almost perfectly match the reconstructed peak positions and the values of w are up to six digits close to α. Also, the values of w in remaining extremal points are less than α, hence the numerical solution μ satisfies, up to a certain accuracy the continuous optimality conditions (3.8). The sparse spike deconvolution method of [17] is also able to recover the exact number of peaks. The solution moreover suffers from the same model-related inaccuracies such as the reduction of magnitude. However, due to the computations being performed on a fixed grid, the optimality conditions do not match as well as for the semi-discrete method of Algorithm 5.1.

The second numerical experiment was performed for $N = 400$ data points, $h_c = h_f = 0.0025$ and an exact solution $\mu^\dagger$ consisting of six delta peaks. Convolution kernel and noise level were the same as for the first experiment, the regularization parameter was $\alpha = 7$. With the stopping criterion as above, the algorithm took 9874 iterations to compute an approximate solution. The results are depicted in Figure 2. Moreover, the tabulated values for the true solution, the approximate solution and the extremal points of the dual variable can be found in Table 2. In a setup analog to the first experiment, we additionally tested the iterative soft-thresholding procedure for sparse spike deconvolution whose outcome is shown in the same table. However, as the primal-dual gap did not reduce sufficiently, we stopped the procedure after 100 000 iterations. The results show that Algorithm 5.1 was again able to detect the correct number of peaks and that the numerical solution almost perfectly satisfies the continuous optimality conditions. In contrast to that, the results for the sparse spike deconvolution are less accurate and took significantly more iterations to compute.

Finally, in a third experiment we were interested in the impact of the acceleration steps. For this purpose, we repeated the first experiment going through all combinations of peak moving and peak merging. As the

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

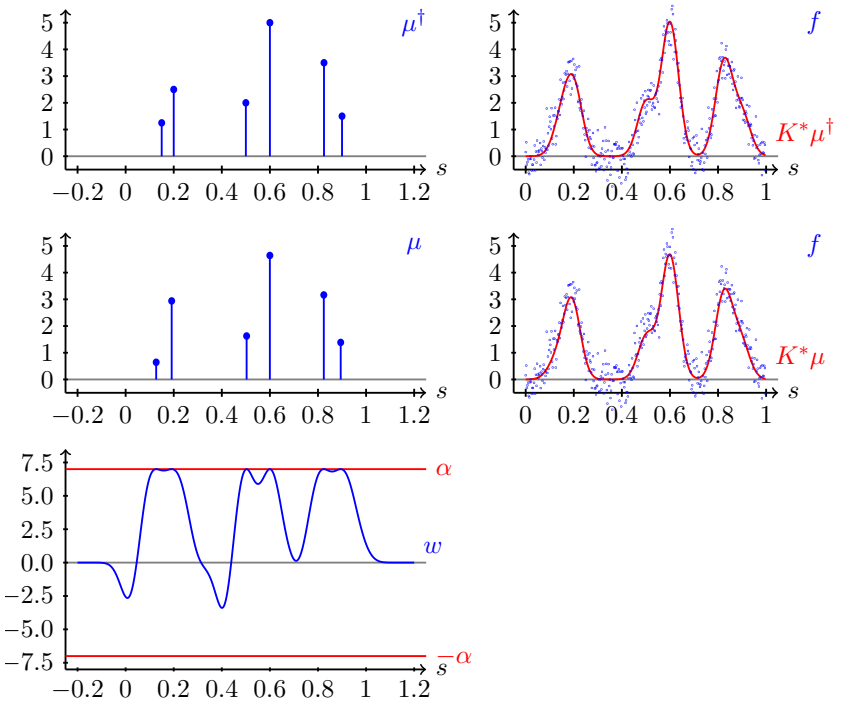


FIGURE 2. Second numerical deconvolution experiment. Upper row: The exact solution $\mu^\dagger$ (left), the exact and noisy data $K^*\mu^\dagger$ and f (right). Second row and third row: the numerical results (also see Fig. 1).

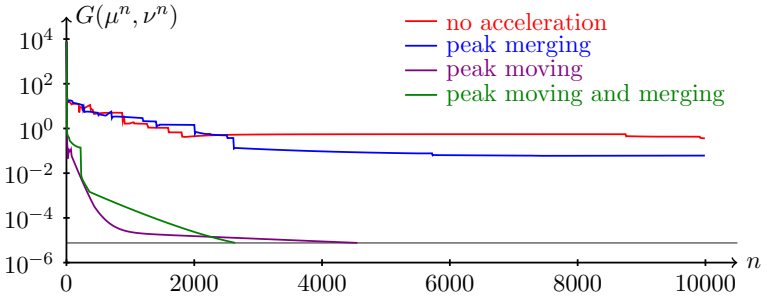


FIGURE 3. Comparison of the evolution of the primal-dual gap $G(\mu^n, \nu^n)$ for different acceleration strategies.

primal-dual gap $G(\mu^n, \nu^n)$ is an upper bound for the functional distance $r(\mu^n)$ (see Lem. 5.5), we observed the evolution of this value with respect to the iteration count, see Figure 3. Note carefully that computation times for each iteration differs for the acceleration strategies. No acceleration is faster than peak moving/merging which is in turn faster than performing both. As a result, we found that no acceleration and pure peak merging is significantly slower than each strategy involving peak moving. In fact, the former two strategies were not able to reduce the primal-dual gap to an acceptable amount within 10 000 iterations. The combination of peak moving and peak merging took the least iterations to reach the stopping criterion.

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

TABLE 2. The values and positions for $\mu^\dagger$, μ and the corresponding extremal points of w for the second deconvolution experiment found by Algorithm 5.1 (top) and iterative soft-thresholding algorithm for sparse spike deconvolution (bottom), respectively (also see Tab. 1).

Successive peak insertion and thresholding					
$\mu^\dagger = \sum_i u_i^\dagger \delta_{x_i^\dagger}$		$\mu = \sum_i u_i \delta_{x_i}$		$w = -K(K^*\mu - f)$	
$x_i^\dagger$	$u_i^\dagger$	x_i	u_i	x_{ex}	$w(x_{\mathrm{ex}})$
0.150000	1.250000	0.127098	0.643334	0.127098	7.000000
0.200000	2.500000	0.191446	2.940414	0.191446	6.999999
0.500000	2.000000	0.503140	1.627235	0.503140	7.000000
0.600000	5.000000	0.599817	4.642194	0.599817	7.000000
0.825000	3.500000	0.824192	3.164055	0.824192	7.000000
0.900000	1.500000	0.894472	1.386177	0.894472	7.000000

Iterative soft-thresholding for sparse spike deconvolution					
$\mu^\dagger = \sum_i u_i^\dagger \delta_{x_i^\dagger}$		$\mu = \sum_i u_i \delta_{x_i}$		$w = -K(K^*\mu - f)$	
$x_i^\dagger$	$u_i^\dagger$	x_i	u_i	x_{ex}	$w(x_{\mathrm{ex}})$
0.150000	1.250000	0.126459	0.625429	0.123647	7.031196
0.200000	2.500000	0.191219	2.959970	0.193652	6.826450
0.500000	2.000000	0.502885	1.620953	0.503216	7.060665
0.600000	5.000000	0.599731	4.651858	0.600533	6.852985
0.825000	3.500000	0.824316	3.173683	0.823512	6.864535
0.900000	1.500000	0.894561	1.377846	0.896583	7.028022

7. Summary and conclusions

Tikhonov regularization for vector-valued finite Radon measures as the solutions of ill-posed linear inverse problems can be put into a proper functional analytic framework which ensures existence, uniqueness (where applicable) and desirable 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. It can easily be amended by various acceleration techniques. As the proposed algorithm produces delta-peak iterates, it is in particular well-suited for the recovery of sparse solutions. Furthermore, the structure of the algorithm allows computations directly on the space of finite Radon measures with only the data space being discretized. However, as the algorithm 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 deconvolution in a reasonable number of iterations; even in the presence of considerable noise. The standard alternative of iterative soft-thresholding on a fixed grid is generally less accurate and needs considerably more iterations.

Appendix A. Weak* density in the space of measures

For the sake of completeness, we like to give a proof of the density results claimed in Remark 2.1.

(i) The density of the set $\{w\mathcal{L}^d \ : \ w \in L^2(\Omega, \mathbb{R}^m)\}$ 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.

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

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

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

for each $u \in \mathcal{C}_0(\Omega, \mathbb{R}^m)$.

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 $|x - y|_2 \leq \varepsilon$. 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

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

Hence $\|u - M_\varepsilon u\|_\infty \to 0$ as $\varepsilon \to 0$. Thus by (A.1) and the continuity,

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

for each $u \in \mathcal{C}(\Omega, \mathbb{R}^m)$ with compact support. This implies the assertion $w^\varepsilon \mathcal{L}^d \rightharpoonup^* \mu$ since convergence was checked against a strongly dense (or norm dense) subset of $\mathcal{C}_0(\Omega, \mathbb{R}^m)$.

(ii) Denote by $Q_x^i = 2^{-i}]-\frac{1}{2}, \frac{1}{2}]^d + x$ for $x \in \mathbb{R}^d$ and $i \in \mathbb{N}$. Since the claim is trivial for $\mu = 0$, pick $\mu \in \mathcal{M}(\Omega, \mathbb{R}^m)$, $\mu \neq 0$ and define the sequence

$$\mu^i = \sum_{x \in 2^{-i}\mathbb{Z}^d \cap \Omega} \mu(Q_x^i \cap \Omega) \delta_x.$$

Obviously, each μ^i is a finite linear combination of δ-peaks with $\|\mu^i\|_\mathcal{M} \leq \|\mu\|_\mathcal{M}$. Choose $u \in \mathcal{C}(\mathbb{R}^d, \mathbb{R}^m)$ with compact support in Ω. We may find, by uniform continuity of u, an $i_0 \in \mathbb{N}$ such that for each $i \geq i_0$, $x \in 2^{-i}\mathbb{Z}^d$ and $y \in Q_x^i$ it follows that $|u(y) - u(x)|_* \leq \varepsilon / \|\mu\|_\mathcal{M}$. Hence,

$$\begin{aligned} \left| \int_\Omega u\, \mathrm{d}\mu - \int_\Omega u\, \mathrm{d}\mu^i \right| &\leq \sum_{x \in 2^{-i}\mathbb{Z}^d} \left| \int_{Q_x^i \cap \Omega} u(y) - u(x)\, \mathrm{d}\mu(y) \right| \\ &\leq \sum_{x \in 2^{-i}\mathbb{Z}^d} \int_{Q_x^i \cap \Omega} |u(y) - u(x)|_*\, \mathrm{d}|\mu| \leq \varepsilon. \end{aligned}$$

This implies $\mu^i \rightharpoonup^* \mu$ as convergence was tested against a set which is strongly dense in $\mathcal{C}_0(\Omega, \mathbb{R}^m)$.

Acknowledgements. Kristian Bredies acknowledges support by the Austrian Science Fund (FWF) special research grant SFB F32 "Mathematical Optimization and Applications in Biomedical Sciences".

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 (2000).
[3] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2** (2009) 183–202.
[4] T. Bonesky, K.S. Kazimierski, P. Maass, F. Schöpfer and T. Schuster, Minimization of Tikhonov functionals in Banach spaces. *Abstr. Appl. Anal.* (2008) 192679.

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

[5] K. Bredies and D.A. Lorenz, Iterated hard shrinkage for minimization problems with sparsity constraints. *SIAM J. Sci. Comput.* **30** (2008) 657–683.
[6] K. Bredies and D.A. Lorenz, Linear convergence of iterative soft-thresholding. *J. Fourier Anal. Appl.* **14** (2008) 813–837.
[7] K. Bredies, D.A. Lorenz and P. Maass, A generalized conditional gradient method and its connection to an iterative shrinkage method. *Comput. Optim. Appl.* **42** (2009) 173–193.
[8] 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*, edited by H.-G. Bock *et al.*, Springer (2010) 287–292.
[9] M. Burger and S. Osher, Convergence rates of convex variational regularization. *Inverse Prob.* **20** (2004) 1411–1421.
[10] E.J. Candès, J.K. Romberg and T. Tao, Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.* **59** (2006) 1207–1223.
[11] C. Clason and K. Kunisch, A duality-based approach to elliptic control problems in non-reflexive Banach spaces. *ESAIM: COCV* **17** (2011) 243–266.
[12] P.L. Combettes and V.R. Wajs, Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.* **4** (2005) 1168–1200.
[13] J.B. Conway, *A course in functional analysis*. Springer (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) 1413–1457.
[15] D.L. Donoho, Compressed sensing. *IEEE Trans. Inf. Theory* **52** (2006) 1289–1306.
[16] D.L. Donoho, M. Elad and V.N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inf. Theory* **52** (2006) 6–18.
[17] C. Dossal and S. Mallat, *Sparse spike deconvolution with minimum scale*, in *Proc. of SPARS'05* (2005).
[18] N. Dunford and J.T. Schwartz, *Linear Operators. I. General Theory*. Interscience Publishers (1958).
[19] B. Efron, T. Hastie, I. Johnstone and R. Tibshirani, Least angle regression. *Ann. Statist.* **32** (2004) 407–499.
[20] I. Ekeland and R. Temam, *Convex analysis and variational problems*. North-Holland (1976).
[21] H.W. Engl and G. Landl, Convergence rates for maximum entropy regularization. *SIAM J. Numer. Anal.* **30** (1993) 1509–1536.
[22] H.W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*. Kluwer Academic Publishers (1996).
[23] 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 J. Sel. Top. Signal Process.* **1** (2007) 586–597.
[24] I. Fonseca and G. Leoni, *Modern methods in the calculus of variations: L^p spaces*. Springer (2007).
[25] M. Fornasier and H. Rauhut, Recovery algorithms for vector valued data with joint sparsity constraints. *SIAM J. Numer. Anal.* **46** (2008) 577–613.
[26] J.-J. Fuchs, On sparse representations in arbitrary redundant bases. *IEEE Trans. Inf. Theory.* **50** (2004) 1341–1344.
[27] A.L. Gibbs and F.E. Su, On choosing and bounding probability metrics. *Int. Stat. Rev.* **70** (2002) 419–435.
[28] M. Grasmair, M. Haltmeier and O. Scherzer, *Sparse regularization with ℓ^q penalty term. Inverse Prob.* **24** (2008) 055020.
[29] R. Griesse and D.A. Lorenz, A semismooth Newton method for Tikhonov functionals with sparsity constraints. *Inverse Prob.* **24** (2008) 035007.
[30] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*. Pitman Publishing Limited (1985).
[31] T. Hein, Tikhonov regularization in Banach spaces – improved convergence rates results. *Inverse Prob.* **25** (2009) 035002.
[32] B. Hofmann, B. Kaltenbacher, C. Pöschl and O. Scherzer, A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Prob.* **23** (2007) 987–1010.
[33] L. Hörmander, *The Analysis of Linear Partial Differential Operators I*. Springer-Verlag (1990).
[34] V.K. Ivanov, V.V. Vasin and V.P. Tanana, *Theory of linear ill-posed problems and its applications*, 2nd edition. Inverse and Ill-posed Problems Series, VSP, Utrecht (2002).
[35] H. Lee, A. Battle, R. Raina and A.Y. Ng, Efficient sparse coding algorithms, in *Advances in Neural Information Processing Systems*, edited by B. Schölkopf, J. Platt and T. Hoffman. MIT Press **19** (2007) 801–808.
[36] J. Lindenstrauss and L. Tzafriri, *Classical Banach Spaces II. Function Spaces*. Springer (1979).
[37] D.A. Lorenz, Convergence rates and source conditions for Tikhonov regularization with sparsity constraints. *J. Inverse Ill-Posed Probl.* **16** (2008) 463–478.
[38] D.A. Lorenz and D. Trede, Optimal convergence rates for Tikhonov regularization in Besov scales. *Inverse Prob.* **24** (2008) 055010.
[39] D.A. Lorenz and D. Trede, Greedy deconvolution of point-like objects, in *Proc. of SPARS'09* (2009).
[40] Y. Mao, B. Dong and S. Osher, A nonlinear PDE-based method for sparse deconvolution. *Multiscale Model. Simul.* **8** (2010) 965–976.
[41] 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) 1841–1854.
[42] Y.E. Nesterov, A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Soviet Math. Dokl.* **27** (1983) 372–376.
[43] A. Neubauer, On enhanced convergence rates for Tikhonov regularization of nonlinear ill-posed problems in Banach spaces. *Inverse Prob.* 25 (2009) 065009.

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press

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

[45] O. Scherzer and B. Walch, Sparsity regularization for Radon measures, in *Scale Space and Variational Methods in Computer Vision*, edited by X.-C. Tai, K. Morken, M. Lysaker and K.-A. Lie. Springer-Verlag (2009) 452–463.

[46] G. Stadler, Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices. *Comput. Optim. Appl.* **44** (2009) 159–181.

[47] G. Stampacchia, Le problème de Dirichlet pour les équations elliptiques du second ordre à coefficients discontinus. *Ann. Inst. Fourier (Grenoble)* **15** (1965) 189–258.

[48] A.S. Stern, D.L. Donoho and J.C. Hoch, NMR data processing using iterative thresholding and minimum l^1-norm reconstruction. *J. Magn. Reson.* **188** (2007) 295–300.

[49] A.N. Tikhonov, A.S. Leonov and A.G. Yagola, *Nonlinear ill-posed problems* **1**. Chapman & Hall (1998).

[50] Z.B. Xu and G.F. Roach, Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces. *J. Math. Anal. Appl.* **157** (1991) 189–210.

[51] C. Zălinescu, *Convex analysis in general vector spaces*. World Scientific (2002).

[52] E. Zeidler, *Nonlinear Functional Analysis and its Applications* III. Springer-Verlag (1985).

https://doi.org/10.1051/cocv/2011205 Published online by Cambridge University Press