Research Network FWF S105

Photoacoustic Imaging in Medicine and Biology

http://pai.uibk.ac.at

Sparsity in Inverse Geophysical Problems

M. Grasmair, M. Haltmeier
and O. Scherzer

December 2009

PAI Report No. 21

Der Wissenschaftsfonds.

Sparsity in Inverse Geophysical Problems

December 1, 2009

Markus Grasmair
University of Vienna
Computational Science Center
Vienna, Austria

Markus Haltmeier
University of Vienna
Computational Science Center
Vienna, Austria

Otmar Scherzer
University of Vienna
Computational Science Center
Vienna, Austria

Abstract. Many geophysical imaging problems are ill-posed in the sense that the solution does not depend continuously on the measured data. Therefore their solutions cannot be computed directly, but instead require the application of regularization. Standard regularization methods find approximate solutions with small L^2 norm. In contrast, sparsity regularization yields approximate solutions that have only a small number of non-vanishing coefficients with respect to a prescribed set of basis elements. Recent results demonstrate that these sparse solutions often much better represent real objects than solutions with small L^2 norm.

In this survey we review recent mathematical results for sparsity regularization. As an application of our theoretical results we consider synthetic focusing in Ground Penetrating Radar, which is a paradigm in inverse geophysical problems.

Contents

1 Introduction

In a plethora of industrial problems one aims at estimating the properties of a physical object from observed data. Often the relation between the physical object and the data can be modeled sufficiently well by a linear equation

$$\mathbf{A}\, u = v \,, \tag{1}$$

where u is a representation of the object in some Hilbert space U, and v a representation of the measurement data, again in a Hilbert space V. Because the operator $\mathbf{A}\colon U \to V$ in general is continuous, the relationship (1) allows one to easily compute data v from the properties of the object u, provided they are known. This is the so called *forward problem*. In many practical applications, however, one is interested in the *inverse problem* of estimating the quantity u from measured data v. A typical feature of inverse problems is that the solution of (1) is very sensitive to perturbations in v. Because in practical applications only an approximation v^δ of the true data v is given, the direct solution of equation (1) by applying the inverse operator is therefore not advisable (see [17, 43]).

By incorporating a–priori information about the exact solution, regularization methods allow to calculate a reliable approximation of u from the observed data v^δ. In this paper we are especially interested in sparsity regularization, where the a–priori information is that the true solution u is sparse in the sense that only few coefficients $\langle u, \phi_\lambda \rangle$ with respect to some prescribed basis $(\phi_\lambda)_{\lambda\in\Lambda}$ are non-vanishing. In the discrete setting of compressed sensing it has recently been shown that sparse solutions can be found by minimizing the ℓ^1-norm of the coefficients $\langle u, \phi_\lambda \rangle$, see [8, 15]. Minimization of the ℓ^1 norm for finding a sparse solutions has however been proposed and studied much earlier for certain geophysical inverse problems (see [9, 33, 38, 42]).

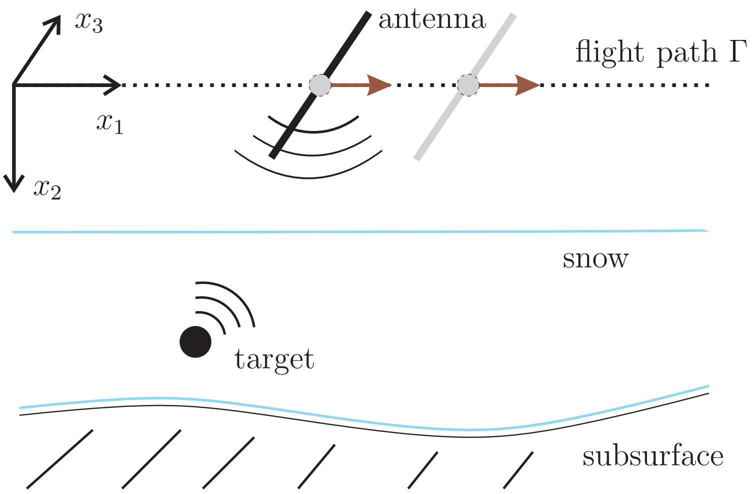

Figure 1: **Collecting GPR data from a flying helicopter.** At each position on the flight path Γ, the antenna emits a short radar pulse. The radar waves get reflected, and the scattered signals are collected in radargrams.

Case Example: Ground Penetrating Radar

As a case example of a geophysical inverse problem we consider Ground Penetrating Radar (GPR), which aims at finding buried objects by measuring reflected radar signals [12]. The reflected signals are detected in zero offset mode (emitting and detecting antenna are at the same position) and used to estimate the reflecting objects. Our interest in GPR has been raised by the possibility of locating avalanche victims by means of a GPR system mounted on a flying helicopter [20, 27]. The basic principle of collecting GPR data from a helicopter is shown in Figure 1.

In Subsection 5.1 below we will show that the imaging problem in GPR reduces to solving the equation (1), with $\mathbf{A}$ being the circular Radon transform. The inversion of the circular Radon transform also arises in several other up-to-date imaging modalities, such as in SONAR, seismic imaging, ultrasound tomography, and photo-/thermo-acoustic tomography (see, e.g., [1, 2, 19, 31, 37, 39, 43, 45] and the reference therein).

2 Variational Regularization Methods

Let U and V be Hilbert spaces and let $\mathbf{A}\colon U \to V$ be a bounded linear operator with unbounded inverse. Then the problem of solving the operator equation

$$\mathbf{A}\, u = v$$

is ill-posed. In order to (approximately) solve this equation in a stable way, it is therefore necessary to introduce some a–priori knowledge about the solution u, which can be expressed via smallness of some *regularization functional* $\mathcal{R}\colon U \to [0,+\infty]$. In classical regularization theory one assumes that the possible solutions have a small energy in some Hilbert space norm—typically an L^2 or H^1-norm is used—and defines $\mathcal{R}$ as the square of this norm. In contrast, we consider below the situation of sparsity constraints, where we assume that the possible solutions have a sparse expansion with respect to a given basis.

We denote by $u^\dagger$ any $\mathcal{R}$-minimizing solution of the equation $\mathbf{A}\, u = v$, provided that it exists, that is,

$$u^\dagger \in \arg\min\bigl\{\mathcal{R}(u) : \mathbf{A}\, u = v\bigr\}\ .$$

In applications, it is to be expected that the measurements v we obtain are disturbed by noise. That is, we are not able to measure the true data v, but only have some noisy measurements v^δ available. In this case, solving the constrained minimization problem $\mathcal{R}(u) \to \min$ subject to $\mathbf{A}\, u = v^\delta$ is not suitable, because the ill-posedness of the equation will lead to unreliable results. Even more, in the worst case it can happen that v^δ is not contained in the range of $\mathbf{A}$, and thus the equation $\mathbf{A}\, u = v^\delta$ has no solution at all. Thus it is necessary to restrict ourselves to solving the given equation only approximately.

We consider three methods for the approximate solution, all of which require knowledge about, or at least some estimate of, the noise level $\delta := \|v - v^\delta\|$.

Residual method: Fix $\tau \geq 1$ and solve the constrained minimization problem

$$\mathcal{R}(u) \to \min \qquad \text{subject to } \|\mathbf{A}\, u - v^\delta\| \leq \tau\delta\ . \tag{2}$$

Tikhonov regularization with discrepancy principle: Fix $\tau \geq 1$ and minimize the *Tikhonov functional*

$$\mathcal{T}_{\alpha,v^\delta}(u) := \|\mathbf{A}\, u - v^\delta\|^2 + \alpha\mathcal{R}(u)\,, \tag{3}$$

where $\alpha > 0$ is chosen in such a way that Morozov's discrepancy principle is satisfied, that is, $\|\mathbf{A}\, u_\alpha^\delta - v^\delta\| = \tau\delta$ with $u_\alpha^\delta \in \arg\min_u \mathcal{T}_{\alpha,v^\delta}(u)$.

Tikhonov regularization with a–priori parameter choice: Fix $C > 0$ and minimize the Tikhonov functional (3) with a parameter choice

$$\alpha = C\delta\ . \tag{4}$$

The residual method aims for the minimization of the penalty term $\mathcal{R}$ over all elements u that generate approximations of the given noisy data v^δ; the size of the permitted defect is dictated by the assumed noise level δ. In particular, the true solution $u^\dagger$ is guaranteed to be among the feasible elements in the minimization problem (2). The additional parameter $\tau \geq 1$ allows for some incertitude concerning the precise noise level; if τ is strictly greater than 1, an underestimation of the noise would still yield a reasonable result.

If the regularization functional $\mathcal{R}$ is convex, the residual method can be shown to be equivalent to Tikhonov regularization with a parameter choice according to Morozov's discrepancy principle, provided the size of the signal is larger than the noise level, that is, the signal-to-noise ratio is larger than τ. In this case, the regularization parameter α in (3) plays the role of a Lagrange parameter for the solution of the constrained minimization problem (2). This equivalence result is summarized in the following theorem (see [30, Thms. 3.5.2, 3.5.5]):

Theorem 2.1. *Assume that the operator* $\mathbf{A}\colon U \to V$ *is linear and has dense range and that the regularization term* $\mathcal{R}$ *is convex. In addition assume that* $\mathcal{R}(u) = 0$ *if and only if* $u = 0$*. Then the residual method and Tikhonov regularization with an a–posteriori parameter choice by means of the discrepancy principle are equivalent in the following sense:*

Let $v^\delta \in V$ *and* $\delta > 0$ *satisfy* $\|v^\delta\| > \tau\delta$*. Then* u^δ *solves the constrained problem (2), if and only if* $\|\mathbf{A}\, u^\delta - v^\delta\| = \tau\delta$ *and there exists some* $\alpha > 0$ *such that* u^δ *minimizes the Tikhnonov functional (3).*

In order to show that the methods introduced above are indeed regularizing, three properties have to be necessarily satisfied, namely existence, stability, and convergence. In addition, convergence rates can be used to quantify the quality of the method:

- *Existence:* For each regularization parameter $\alpha > 0$ and every $v^\delta \in V$ the regularization functional $\mathcal{T}_{\alpha,v^\delta}$ attains its minimum. Similarly, the minimization problem (2) has a solution.
- *Stability* is required to ensure that, for fixed noise level δ, the regularized solutions depend continuously on the data v^δ.
- *Convergence* ensures that the regularized solution converge to $u^\dagger$ as the noise level decreases to zero.
- *Convergence rates* provide an estimate of the difference between the minimizers of the regularization functional and $u^\dagger$.

Typically, convergence rates are formulated in terms of the *Bregman distance* (see [6, 29, 41, 43]), which, for a convex and differentiable regularization term $\mathcal{R}$ with subdifferential $\partial\mathcal{R}$ and $\xi \in \partial\mathcal{R}(u^\dagger)$, is defined as

$$\mathcal{D}(u, u^\dagger) = \mathcal{R}(u) - \mathcal{R}(u^\dagger) - \langle \xi, u - u^\dagger \rangle \ .$$

That is, $\mathcal{D}(u, u^\dagger)$ measures the distance between the tangent and the convex function $\mathcal{R}$. In general, convergence with respect to the Bregman distance does not imply convergence with respect to the norm, strongly reducing the significance of the derived rates. In the setting of sparse regularization to be introduced below, however, it is possible to derive convergence rates with respect to the norm on U.

3 Sparse Regularization

In the following we concentrate on sparsity promoting regularization methods. To that end, we assume that $(\phi_\lambda)_{\lambda\in\Lambda}$ is an orthonormal basis of the Hilbert space U, for instance a wavelet or Fourier basis. For $u \in U$, we denote by

$$\operatorname{supp}(u) := \{\lambda \in \Lambda : \langle \phi_\lambda, u \rangle \neq 0\}$$

the *support* of u with respect to the basis $\big(\phi_\lambda\big)_{\lambda\in\Lambda}$. If $|\operatorname{supp}(u)| \leq s$ for some $s \in \mathbb{N}$, then the element u is called s*-sparse*. It is called *sparse*, if it is s-sparse for some $s \in \mathbb{N}$, that is, $|\operatorname{supp}(u)| < \infty$. Given weights w_λ, $\lambda \in \Lambda$, bounded below by some constant $w_{\min} > 0$, we define for $0 < q \leq 2$ the ℓ^q-regularization functional

$\mathcal{R}_q \colon U \to \mathbb{R} \cup \{\infty\}$,

$$\mathcal{R}_q(u) := \sum_{\lambda\in\Lambda} w_\lambda |\langle \phi_\lambda, u \rangle|^q \ .$$

If $q = 2$, then the regularization functional is simply the weighted squared Hilbert space norm on U.

If q is smaller than 2, small coefficients $\langle \phi_\lambda, u \rangle$ are penalized comparatively stronger, while the penalization of large coefficients becomes less pronounced. As a consequence, the reconstructions resulting by applying any of the above introduced regularization methods will exhibit a small number of significant coefficients, while most of the coefficients will be close to zero. These sparsity enhancing properties of ℓ^q-regularization become more pronounced as the parameter q decreases. If we choose q at most 1, then the reconstructions are necessarily sparse in the above, strict sense, that is, the number of non-zero coefficients is at most finite (see [21]):

Proposition 3.1. *Let $q \leq 1$, $\alpha > 0$, $v^\delta \in V$. Then every minimizer of the Tikhonov functional $\mathcal{T}_{\alpha,v^\delta}$ with regularization term $\mathcal{R}_q$ is sparse.*

There are compelling reasons for using an exponent $q \geq 1$ in applications, as this choice entails the convexity of the ensuing regularization functionals. In contrast, a choice $q < 1$ leads to non-convex minimization problems and, as a consequence, to numerical difficulties in their minimization. In the convex case $q \geq 1$, there are several possible strategies for computing the minimizers of regularization functional $\mathcal{T}_{\alpha,v^\delta}$. Below, in Section 4, we will consider two different, iterative methods: an Iterative Thresholding Algorithm for regularization with a–priori parameter choice and $1 \leq q \leq 2$ [14], and a log-barrier method for Tikhonov regularization with an a–posteriori parameter choice by the discrepancy principle in the case $q = 1$ [7]. Iterative thresholding algorithms have also been studied for non-convex situations, but there the convergence to global minima has not yet been proven [5].

3.1 Convex Regularization

We now turn to the study of the theoretical properties of ℓ^q type regularization methods with $q \geq 1$ and study the questions of existence, stability, convergence, and convergence rates. In order to be able to take advantage of the equivalence result Theorem 2.1, we assume in the following that the operator $\mathbf{A} \colon U \to V$ has dense range.

The question of existence is easily answered [23, 25]:

Proposition 3.2 (Existence)**.** *For every $\alpha > 0$ and $v^\delta \in V$ the functional $\mathcal{T}_{\alpha,v^\delta}$ has a minimizer in U. Similarly, the problem of minimizing $\mathcal{R}_q(u)$ subject to the constraint $\|\mathbf{A}\, u - v^\delta\| \leq \tau\delta$ admits a solution in U.*

Though the previous lemma states the existence of minimizers for all $q \geq 1$, there is a difference between the cases $q = 1$ and $q > 1$. In the latter case, the regularization functional $\mathcal{T}_{\alpha,v^\delta}$ is strictly convex, which implies that the minimizer must be unique. For $q = 1$, the regularization functional is still convex, but the strict convexity holds only, if the operator $\mathbf{A}$ is injective. Thus it can happen that we do not obtain a single approximate solution, but a whole (convex and closed) set of minimizers. Because of this possible non-uniqueness, the stability and convergence results have to be formulated in terms of subsequential convergence.

Also, we have to differentiate between a–priori and a–posteriori parameter selection methods. In the latter case, the stability and convergence results can be formulated solely in terms of the noise level δ. In the case of an a–priori parameter choice, it is in addition necessary to take into account the actual choice of α in dependence of δ. For the following results we refer to [25, 34].

Proposition 3.3 (Stability)**.** *Let $\delta > 0$ be fixed and let $v_k \to v^\delta$. Consider one of the following settings:*

Residual method: *Let* $u_k \in U$ *be solutions of the residual method with data* v_k *and noise level* δ*.*

Discrepancy principle: *Let* $u_k \in U$ *be solutions of Tikhonov regularization with data* v_k *and an a–posteriori parameter choice according to the discrepancy principle for noise level* δ*.*

A–priori parameter choice: *Let* $\alpha > 0$ *be fixed, and let* $u_k \in U$ *be solutions of Tikhonov regularization with data* v_k *and regularization parameter* α*.*

Then the sequence $(u_k)_{k\in\mathbb{N}}$ *has a subsequence converging to a regularized solution* u^δ *obtained with data* v^δ *and the same regularization method. If* u^δ *is unique, then the whole sequence* $(u_k)_{k\in\mathbb{N}}$ *converges to* u^δ*.*

Proposition 3.4 (Convergence)**.** *Let* $\delta_k \to 0$ *and let* $v_k \in V$ *satisfy*

$$\|v_k - v\| \leq \delta_k \,.$$

Assume that there exists $u \in U$ *with* $\mathbf{A}\, u = v$ *and* $\mathcal{R}_q(u) < +\infty$*. Consider one of the following settings:*

Residual method: *Let* $u_k \in U$ *be solutions of the residual method with data* v_k *and noise level* δ_k*.*

Discrepancy principle: *Let* $u_k \in U$ *be solutions of Tikhonov regularization with data* v_k *and an a–posteriori parameter choice according to the discrepancy principle with noise level* δ_k*.*

A–priori parameter choice: *Let* $\alpha_k > 0$ *satisfy* $\alpha_k \to 0$ *and* $\delta_k^2/\alpha_k \to 0$*, and let* $u_k \in U$ *be solutions of Tikhonov regularization with data* v_k *and regularization parameter* α_k*.*

Then the sequence $(u_k)_{k\in\mathbb{N}}$ *has a subsequence converging to an* $\mathcal{R}_q$*-minimizing solution* $u^\dagger$ *of the equation* $\mathbf{A}\, u = v$*. If* $u^\dagger$ *is unique, then the whole sequence* $(u_k)_{k\in\mathbb{N}}$ *converges to* $u^\dagger$*.*

Note that the previous result in particular implies that an $\mathcal{R}_q$-minimizing solution $u^\dagger$ of $\mathbf{A}\, u = v$ indeed exists. Also, the uniqueness of $u^\dagger$ is trivial in the case $q > 1$, as then the functional $\mathcal{R}_q$ is strictly convex. Thus we obtain in this situation indeed convergence of the whole sequence $(u_k)_{k\in\mathbb{N}}$.

Though we know now that approximative solutions indeed converge to true solutions of the considered equation as the noise level decreases to zero, we have obtained no estimate for the speed of the convergence. Indeed, in general situations the convergence can be arbitrarily slow. If, however, the $\mathcal{R}_q$-minimizing solution $u^\dagger$ satisfies a so-called *source condition*, then we can obtain sufficiently good convergence rates in the strictly convex case $q > 1$. If, in addition, the solution $u^\dagger$ is sparse and the operator $\mathbf{A}$ is invertible on the support of $u^\dagger$, then the convergence rates improve further.

Before stating the convergence rates results, we recall the definition of the source condition and its relation to the well-known Karush–Kuhn–Tucker condition used in convex optimization.

Definition 3.5. The $\mathcal{R}_q$-minimizing solution $u^\dagger$ of the equation $\mathbf{A}\, u = v$ satisfies the source condition, if there exists $\xi \in V$ such that $\mathbf{A}^*\, \xi \in \partial \mathcal{R}_q(u^\dagger)$. Here $\partial \mathcal{R}_q(u^\dagger)$ denotes the subdifferential of the function $\mathcal{R}_q$ at $u^\dagger$, and $\mathbf{A}^* \colon V \to U$ is the adjoint of $\mathbf{A}$.

In other words, if $q > 1$ we have

$$\langle \xi, \mathbf{A}\, \phi_\lambda \rangle = q \operatorname{sign}(\langle u^\dagger, \phi_\lambda \rangle) |\langle u^\dagger, \phi_\lambda \rangle|^{q-1}\,, \quad \lambda \in \Lambda\,,$$

and if $q = 1$ we have

$$\begin{aligned} \langle \xi, \mathbf{A}\, \phi_\lambda \rangle &= \operatorname{sign}(\langle u^\dagger, \phi_\lambda \rangle) && \text{if } \lambda \in \operatorname{supp}(u^\dagger)\,, \\ \langle \xi, \mathbf{A}\, \phi_\lambda \rangle &\in [-1, +1] && \text{if } \lambda \notin \operatorname{supp}(u^\dagger)\,. \end{aligned}$$

The conditions $\mathbf{A}^*\, \xi \in \partial \mathcal{R}_q(u^\dagger)$ for some $\xi \in V$ and $\mathbf{A}\, u^\dagger = v$ are nothing more than the Karush–Kuhn–Tucker conditions for the constrained minimization problem

$$\mathcal{R}_q(u) \to \min \qquad \text{subject to } \mathbf{A}\, u = v\,.$$

In particular, it follows that $\tilde{u} \in U$ is an $\mathcal{R}_q$-minimizing solution of the equation $\mathbf{A}\, u = v$ whenever $\tilde{u}$ satisfies the equation $\mathbf{A}\, \tilde{u} = v$ and we have $\operatorname{ran} \mathbf{A}^* \cap \partial \mathcal{R}_q(\tilde{u}) \neq \emptyset$ [16, Proposition 4.1].

The following convergence rates result can be found in [25, 34]. It is based on results concerning convergence rates with respect to the Bregman distance (see [6]) and the fact that, for ℓ^q-regularization, the norm can be bounded from above, locally, by the Bregman distance.

Proposition 3.6. *Let $1 < q \leq 2$ and assume that $u^\dagger$ satisfies the source condition. Denote, for $v^\delta \in V$ satisfying $\|v^\delta - v\| \leq \delta$, by $u^\delta := u(v^\delta)$ the solution with data v^δ of either the residual method, or Tikhonov regularization with Morozov's discrepancy principle, or Tikhonov regularization with an a–priori parameter choice $\alpha = C\delta$ for some fixed $C > 0$. Then*

$$\|u^\delta - u^\dagger\| = O(\sqrt{\delta})\,.$$

In the case of an a–priori parameter choice, we additionally have that

$$\|\mathbf{A}\, u^\delta - v\| = O(\delta)\,.$$

The convergence rates provide (asymptotic) estimates of the accuracy of the approximative solution in dependence of the noise level δ. Therefore the optimization of the order of convergence is an important question in the field of inverse problems.

In the case of Tikhonov regularization with a–priori parameter choice, the rates can indeed be improved, if the stronger source condition $\mathbf{A}^*\, \mathbf{A}\, \eta \in \partial \mathcal{R}_q(u^\dagger)$ for some $\eta \in U$ holds. Then one obtains with a parameter choice

$\alpha = C\delta^{2/3}$ a rate of order $O(\delta^{2/3})$ (see [26, 41]). For quadratic Tikhonov regularization it has been shown that this rate is the best possible one. That is, except in the trivial case $u^\dagger = 0$, there exists no parameter selection method, neither a–priori nor a–posteriori, that can yield a better rate than $O(\delta^{2/3})$ (see [36]). This saturation result poses a restriction on the quality of reconstructions obtainable with quadratic regularization.

In the non-quadratic case $q < 2$ the situation looks different. If the solution $u^\dagger$ is sparse, then the convergence rates results can be improved beyond the quadratic bound of $O(\delta^{2/3})$. Moreover, they also can be extended to the case $q = 1$. For the improvement of the convergence rates, an additional injectivity condition is needed, which requires the operator $\mathbf{A}$ to be injective on the (finite dimensional) subspace of U spanned by the basis elements ϕ_λ, $\lambda \in \operatorname{supp}(u^\dagger)$. This last condition is trivially satisfied, if the operator $\mathbf{A}$ itself is injective. There exist, however, also interesting situations, where the linear equation $\mathbf{A}\,u = v$ is vastly under-determined, but the restriction of $\mathbf{A}$ to all sufficiently low-dimensional subspaces spanned by the basis elements ϕ_λ is injective. These cases have recently been well studied in the context of *compressed sensing* [8, 15]. The first improved convergence rates have been derived in [23, 25].

Proposition 3.7. *Let $q \geq 1$ and assume that $u^\dagger$ satisfies the source condition. In addition, assume that $u^\dagger$ is sparse and that the restriction of the operator $\mathbf{A}$ to $\operatorname{span}\{\phi_\lambda : \lambda \in \operatorname{supp}(u^\dagger)\}$ is injective.*

Then, with the notation of Proposition 3.6, we have

$$\|u^\delta - u^\dagger\| = O(\delta^{1/q}) \ .$$

The most interesting situation is the case $q = 1$. Here, one obtains a linear convergence of the regularized solutions to $u^\dagger$. That is, the approximative inversion of $\mathbf{A}$ is not only continuous, but in fact Lipschitz continuous; the error in the reconstruction is of the same order as the data error. In addition, the source condition $\mathbf{A}^*\,\xi \in \partial\mathcal{R}_q(u^\dagger)$ in some sense becomes weakest for $q = 1$, because then the subdifferential is set-valued and therefore larger than in the strictly convex case. Moreover, the source condition for $q > 1$ requires that the support of $\mathbf{A}^*\,\xi$ equals the support of $u^\dagger$, which strongly limits the applicability of the convergence rates result.

While Proposition 3.7 concerning convergence rates in the presence of a sparsity assumption and restricted injectivity holds for all $1 \leq q \leq 2$, the rates result without these assumptions, Proposition 3.6, requires that the parameter q is strictly greater than one. The following converse result shows that, at least for Tikhonov regularization with an a–priori parameter choice, a similar relaxation of the assumptions by dropping the requirement of restricted injectivity is not possible for $q = 1$; the assumptions of sparsity and injectivity of $\mathbf{A}$ on $\operatorname{supp}(u^\dagger)$ are not only sufficient but also necessary for obtaining any sensible convergence rates (see [24]).

Proposition 3.8. *Let $q = 1$ and assume that $u^\dagger$ is the* unique *$\mathcal{R}_1$-minimizing solution of the equation $\mathbf{A}\,u = v$. Denote, for $v^\delta \in V$ satisfying $\|v^\delta - v\| \le \delta$, by $u^\delta := u(v^\delta)$ the solution with data v^δ of Tikhonov regularization with an a–priori parameter choice $\alpha = C\delta$ for some fixed $C > 0$. If the obtained data error satisfies*

$$\|\mathbf{A}\,u^\delta - v\| = O(\delta)\,,$$

then $u^\dagger$ is sparse and the source condition holds. In particular, also

$$\|u^\delta - v\| = O(\delta)\;.$$

3.2 Non-convex Regularization

In the following, we will study the properties of ℓ^q regularization with a sub-linear regularization term, that is, $0 < q < 1$. In this situation, the regularization functional is non-convex, leading to both theoretical and numerical challenges. Still, non-convex regularization terms have been considered for applications, because they yield solutions with even more pronounced sparsity patterns than ℓ^1 regularization.

From the theoretical point of view, the lack of convexity prohibits the application of Theorem 2.1, which states that the residual method is equivalent to Tikhonov regularization with Morozov's discrepancy principle. Indeed, it seems that an extension of said result to non-convex regularization functionals has not been treated in the literature so far. Even more, though corresponding results have recently been formulated for the residual method, the question, whether the discrepancy principle yields stable reconstructions, has not yet been answered. For these reasons, we limit the discussion of non-convex regularization methods to the two cases of the residual method and Tikhonov regularization with an a–priori parameter choice. Both methods allow the derivation of basically the same, or at least similar, results as for convex regularization, the main difference being the possible non-uniqueness of the $\mathcal{R}_q$-minimizing solutions of the equation $\mathbf{A}\,u = v$ (see [22, 25, 46]).

Proposition 3.9. *Consider either the residual method or Tikhonov regularization with an a–priori parameter choice. Then Propositions 3.2–3.4 concerning existence, stability, and convergence remain to hold true for $0 < q < 1$.*

Also the convergence rates result in the presence of sparsity, Proposition 3.7, can be generalized to non-convex regularization. The interesting point is that the source condition needed in the convex case apparently is not required any more. Instead, the other conditions of Proposition 3.7, uniqueness and sparsity of $u^\dagger$ and restricted injectivity of $\mathbf{A}$, are already sufficient for obtaining linear convergence (see [21, 25]).

Proposition 3.10. *Let $0 < q < 1$ and assume that $u^\dagger$ is the* unique *$\mathcal{R}_q$-minimizing solution of the equation $\mathbf{A}\,u = v$. Assume moreover that $u^\dagger$ is sparse and that the restriction of the operator $\mathbf{A}$ to $\operatorname{span}\{\phi_\lambda : \lambda \in \operatorname{supp}(u^\dagger)\}$ is injective. Denote, for $v^\delta \in V$ satisfying $\|v^\delta - v\| \le \delta$, by $u^\delta := u(v^\delta)$ the solution with data v^δ of either the residual method or Tikhonov regularization with an a–priori parameter choice $\alpha = C\delta$ for some fixed $C > 0$. Then*

$$\|u^\delta - u^\dagger\| = O(\delta)\ .$$

In the case of Tikhonov regularization with an a–priori parameter choice, we additionally have that

$$\|\mathbf{A}\,u^\delta - v\| = O(\delta)\ .$$

4 Numerical Minimization

4.1 Iterative Thresholding Algorithms

In [14], an iterative algorithm has been analyzed that can be used for minimizing the Tikhonov functional $\mathcal{T}_{\alpha,v^\delta}$ for fixed $\alpha > 0$, that is, for an a–priori parameter choice. To that end, we define for $b > 0$ and $1 \le q \le 2$ the function $F_{b,q} \colon \mathbb{R} \to \mathbb{R}$,

$$F_{b,q}(t) := t + \frac{bq}{2}\operatorname{sign}(t)|t|^{q-1}\ .$$

If $q > 1$, the function $F_{b,q}$ is a one-to-one mapping from $\mathbb{R}$ to $\mathbb{R}$. Thus, it has an inverse $S_{b,q} := (F_{b,q})^{-1} \colon \mathbb{R} \to \mathbb{R}$. In the case $q = 1$ we define

$$S_{b,1}(t) := \begin{cases} t - b/2 & \text{if } t \ge b/2\,, \\ 0 & \text{if } |t| < b/2\,, \\ t + b/2 & \text{if } t \le -b/2\ . \end{cases} \tag{5}$$

Using the functions $S_{b,q}$, we define now, for $\mathbf{b} = (b_\lambda)_{\lambda\in\Lambda} \in \mathbb{R}^\Lambda_{>0}$ and $1 \le q \le 2$, the *Shrinkage Operator* $\mathbf{S}_{\mathbf{b},q} \colon U \to U$,

$$\mathbf{S}_{\mathbf{b},q}(u) := \sum_{\lambda\in\Lambda} S_{b_\lambda,q}\big(\langle u, \phi_\lambda\rangle\big)\phi_\lambda\ . \tag{6}$$

Proposition 4.1. *Let $v^\delta \in V$, $\alpha > 0$, and $1 \le q \le 2$, and denote $\mathbf{w} := (w_\lambda)_{\lambda\in\Lambda}$. Let $\mu > 0$ be such that $\mu\|\mathbf{A}^*\,\mathbf{A}\| < 1$. Choose any $u_0 \in U$ and define inductively*

$$u_{n+1} := \mathbf{S}_{\mu\alpha\mathbf{w},q}\big(u_n + \mu\,\mathbf{A}^*(v^\delta - \mathbf{A}\,u_n)\big)\ . \tag{7}$$

Then the iterates u_n, defined by the iterative thresholding iteration (7), converge to a minimizer of the functional $\mathcal{T}_{\alpha,v^\delta}$ as $n \to \infty$.

The method defined by the iteration (7) can be seen as a forward–backward splitting algorithm for the minimization of $\mathcal{T}_{\alpha,v^\delta}$, the inner update $u \mapsto u + \mu\,\mathbf{A}^*(v^\delta - \mathbf{A}\,u)$ being a gradient descent step for the functional $\|\mathbf{A}\,u - \mathbf{A}\,v^\delta\|^2$, and the shrinkage operator a gradient descent step for $\alpha\mathcal{R}_q$. More details on the application of forward–backward splitting methods to similar problems can, for instance, be found in [10].

4.2 Second Order Cone Programs

In the case of an a–posteriori parameter choice (or the equivalent residual method), the iterative thresholding algorithm (7) cannot be applied directly, as the regularization parameter $\alpha > 0$ is not known in advance. One can show, however, that the required parameter α depends continuously on δ (see [3]). Thus it is possible to find the correct parameter iteratively, starting with some initial guess $\alpha > 0$ and computing some $\hat{u} \in \arg\min_u \mathcal{T}_{\alpha,v^\delta}(u)$. Depending on the size of the residual $\mathbf{A}\,\hat{u} - v^\delta$, one subsequently either increases or decreases α and computes the minimizer of $\mathcal{T}_{\alpha,v^\delta}$ using the new regularization parameter. This procedure of updating α and minimizing $\mathcal{T}_{\alpha,v^\delta}$ is stopped, as soon as the residual satisfies $\|\mathbf{A}\,\hat{u} - v^\delta\| \approx \tau\delta$.

In the important case $q = 1$, a different solution algorithm has been established, which takes advantage on the fact that the constrained minimization problem $\mathcal{R}_1(u) \to \min$ subject to $\|\mathbf{A}\,u - v^\delta\|^2 \le \delta$ can be rewritten as a second-order cone program (SOCP). To that end we introduce an additional variable $a = (a_\lambda)_{\lambda\in\Lambda} \in \ell^2(\Lambda)$ and minimize $\sum_{\lambda\in\Lambda} w_\lambda a_\lambda$ subject to the constraints $a_\lambda \ge |\langle u, \phi_\lambda\rangle|$ for all $\lambda \in \Lambda$ and $\|\mathbf{A}\,u - v^\delta\|^2 \le \tau\delta^2$. The former bound consisting of the two linear constraints $a_\lambda \ge \pm\langle u, \phi_\lambda\rangle$, we arrive at the SOCP

$$\mathcal{S}(u,a) := \sum_{\lambda\in\Lambda} w_\lambda a_\lambda \to \min \qquad \text{subject to} \qquad \begin{aligned} a_\lambda + \langle u, \phi_\lambda\rangle &\ge 0\,,\\ a_\lambda - \langle u, \phi_\lambda\rangle &\ge 0\,,\\ \tau\delta^2 - \|\mathbf{A}\,u - v^\delta\|^2 &\le 0\,. \end{aligned} \tag{8}$$

If the pair (u, a) solves (8), then u is a solution of the residual method.

The solutions of the program (8) can be computed using a log-barrier method, defining for $\eta > 0$ the functional

$$\begin{aligned}\mathcal{S}_\eta(u,a) := \eta\sum_{\lambda\in\Lambda} w_\lambda a_\lambda &- \sum_{\lambda\in\Lambda}\log\bigl(a_\lambda + \langle u,\phi_\lambda\rangle\bigr)\\ &- \sum_{\lambda\in\Lambda}\log\bigl(a_\lambda - \langle u,\phi_\lambda\rangle\bigr) - \log\bigl(\|\mathbf{A}\,u - v^\delta\|^2 - \tau\delta^2\bigr)\,.\end{aligned}$$

As $\eta \to \infty$, the minimizers of $\mathcal{S}_\eta(u,a)$ converge to a solution of (8). Moreover, one can show that the solution (u^δ, a^δ) of (8) and the minimizer $(u^\delta_\eta, a^\delta_\eta)$ of $\mathcal{S}_\eta$ satisfy the relation

$$\mathcal{S}(u^\delta_\eta, a^\delta_\eta) < \mathcal{S}(u^\delta, a^\delta) + (|\Lambda| + 1)/\eta\,, \tag{9}$$

that is, the value of the minimizer of the relaxed problem $\mathcal{S}_\eta$ lies within $(|\Lambda|+1)/\eta$ of the optimal value of the original minimization problem [40].

In order to solve (8), one alternatingly minimizes $\mathcal{S}_\eta$ and increases the parameter η. That is, one chooses some parameter $\mu > 1$ defining the increase of η and starts with $k = 1$ and some initial parameter $\eta^{(1)} > 0$. Then one iteratively computes $(u_k, a_k) \in \arg\min \mathcal{S}_{\eta^{(k)}}$, set $\eta^{(k+1)} := \mu\eta^{(k)}$ and increases k until the value $(|\Lambda|+1)/\eta^{(k)}$ is smaller than some predefined tolerance—according to (9), this implies that also the value $\mathcal{S}(u_k, a_k)$ is within the same tolerance of the actual minimum. For the minimization of $\mathcal{S}_{\eta^{(k)}}$, which has to take place in each iteration step, one can use a Newton method combined with a line search that ensures that one does not leave the domain of $\mathcal{S}_{\eta^{(k)}}$ and that the value of $\mathcal{S}_{\eta^{(k)}}$ actually decreases. More details on the minimization algorithm can be found in [7].

5 Application: Synthetic Focusing in Ground Penetrating Radar

In this section we apply sparsity regularization to data obtained with Ground Penetrating Radar (GPR) mounted on a flying helicopter (see Figure 1). As stated in the introduction, we first write the imaging problem as the inversion of the circular Radon transform.

5.1 Mathematical Model

For simplicity of presentation we ignore polarization effects of the electromagnetic field and assume a small isotropic antenna. In this case, each component of the electromagnetic field $E(\mathbf{x}^{\mathrm{ant}}; \mathbf{x}, t)$ induced by an antenna that is located at $\mathbf{x}^{\mathrm{ant}} \in \mathbb{R}^3$ is described by the scalar wave equation

$$\left(\frac{1}{c(\mathbf{x})^2}\,\partial_t^2 - \Delta\right) E(\mathbf{x}^{\mathrm{ant}}; \mathbf{x}, t) = \delta_{3D}(\mathbf{x} - \mathbf{x}^{\mathrm{ant}}) w_b(t)\,, \quad (\mathbf{x}, t) \in \mathbb{R}^3 \times \mathbb{R}\,. \tag{10}$$

Here δ_{3D} denotes the three dimensional delta distribution, w_b represents the temporal shape of the emitted radar signal (impulse response function of the antenna) with bandwidth b, and $c(\mathbf{x})$ denotes the wave speed.

GPR systems are designed to generate ultrawideband radar signals, where the bandwidth b is approximately equal to the central frequency, and the pulse duration is given by $\tau = 1/b$. Usually, w_b is well approximated by the second derivative of a small Gaussian (Ricker wavelet), see [12]. Figure 2 shows a typical radar signal emitted by a radar antenna at 500 MHz and its Fourier transform.

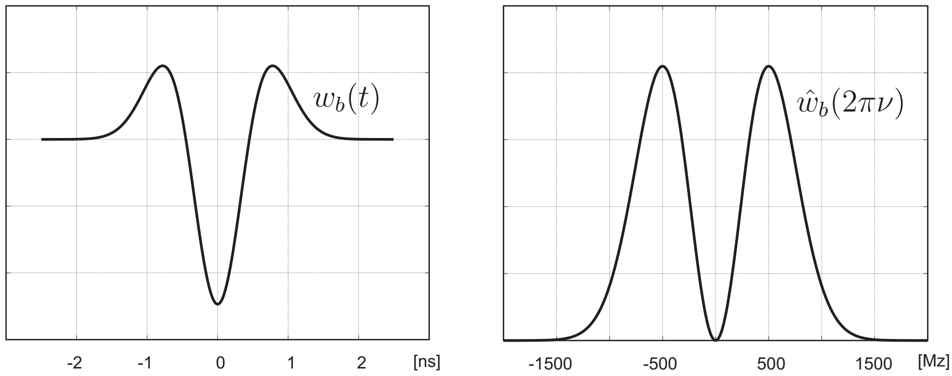

Figure 2: Ricker Wavelet (second derivative of a small Gaussian) with a central frequency of $b = 500$MHz in the time domain (left) and in the frequency domain (right).

Born Approximation

Scattering of the radar signals occurs at discontinuities of the function c. In the sequel, we assume that

$$\frac{1}{c(\mathbf{x})^2} = \frac{1}{c_0^2}\left(1 + u^{3D}(\mathbf{x})\right),$$

where c_0 is assumed to be constant (the light speed) and u^{3D} is a possibly non-smooth function. Moreover, we make the decomposition

$$E(\mathbf{x}^{\text{ant}}; \mathbf{x}, t) = E_0(\mathbf{x}^{\text{ant}}; \mathbf{x}, t) + E_{\text{scat}}(\mathbf{x}^{\text{ant}}; \mathbf{x}, t), \qquad (\mathbf{x}, t) \in \mathbb{R}^3 \times \mathbb{R},$$

where E_0 denotes the incident field (the solution of the wave equation (10) with c replaced by c_0), and E_{scat} is the scattered field.

From (10) it follows that the scattered field satisfies

$$\left(\frac{1}{c_0^2}\,\partial_t^2 - \Delta\right) E_{\text{scat}}(\mathbf{x}^{\text{ant}}; \mathbf{x}, t) = -\frac{u^{3D}(\mathbf{x})}{c_0^2}\frac{\partial^2 E(\mathbf{x}^{\text{ant}}; \mathbf{x}, t)}{\partial t^2}.$$

The Born approximation consist in replacing the total field E in the above equation by the incident field E_0. This results in the approximation $E_{\text{scat}} \simeq E_{\text{Born}}$, where E_{Born} solves the equation

$$\left(\frac{1}{c_0^2}\,\partial_t^2 - \Delta\right) E_{\text{Born}}(\mathbf{x}^{\text{ant}}; \mathbf{x}, t) = -\frac{u^{3D}(\mathbf{x})}{c_0^2}\frac{\partial^2 E_0(\mathbf{x}^{\text{ant}}; \mathbf{x}, t)}{\partial t^2}, \qquad (\mathbf{x}, t) \in \mathbb{R}^3 \times \mathbb{R}. \tag{11}$$

Together with the initial condition $E_{\text{scat}}(\mathbf{x}^{\text{ant}}; \mathbf{x}, t) = 0$ for $t < t_0$, Equation (11) can be solved explicitly via Kirchhoff's formula, see [11, page 692],

$$E_{\text{Born}}(\mathbf{x}^{\text{ant}}; \mathbf{x}, t) = -\frac{1}{4\pi c_0^2}\frac{\partial^2}{\partial t^2}\int_{\mathbb{R}^3} u^{3D}(\mathbf{y})\frac{E_0\left(\mathbf{x}^{\text{ant}}; \mathbf{y}, t - \frac{|\mathbf{x}-\mathbf{y}|}{c_0}\right)}{|\mathbf{x}-\mathbf{y}|}\,d\mathbf{y}.$$

The identity

$$E_0(\mathbf{x}^{\text{ant}};\mathbf{y},t) = -\frac{w_b\big(t-|\mathbf{y}-\mathbf{x}^{\text{ant}}|/c_0\big)}{4\pi|\mathbf{y}-\mathbf{x}^{\text{ant}}|} = -\frac{(w_b *_t \delta_{1D})\big(t-|\mathbf{y}-\mathbf{x}^{\text{ant}}|/c_0\big)}{4\pi|\mathbf{y}-\mathbf{x}^{\text{ant}}|}\,,$$

with δ_{1D} denoting the one dimensional delta distribution, leads to

$$E_{\text{Born}}(\mathbf{x}^{\text{ant}};\mathbf{x},t) = \frac{w_b''(t)}{16\pi^2c_0^2} *_t \int_{\mathbb{R}^3} u^{3D}(\mathbf{y})\,\frac{\delta_{1D}\Big(t-\frac{|\mathbf{y}-\mathbf{x}^{\text{ant}}|}{c_0}-\frac{|\mathbf{x}-\mathbf{y}|}{c_0}\Big)}{|\mathbf{x}-\mathbf{y}|\;|\mathbf{y}-\mathbf{x}^{\text{ant}}|}\,d\mathbf{y}\,. \tag{12}$$

In GPR, the data are measured in zero offset mode, which means that the scattered field is only recorded at location $\mathbf{x} = \mathbf{x}^{\text{ant}}$. In this situation, equation (12) simplifies to

$$E_{\text{Born}}(\mathbf{x}^{\text{ant}};\mathbf{x}^{\text{ant}},t) = \frac{w_b''(t)}{32\pi^2c_0} *_t \int_{\mathbb{R}^3} u^{3D}(\mathbf{y})\,\frac{\delta_{1D}\Big(\frac{c_0t}{2}-|\mathbf{y}-\mathbf{x}^{\text{ant}}|\Big)}{|\mathbf{y}-\mathbf{x}^{\text{ant}}|^2}\,d\mathbf{y}\,,$$

where we made use of the formula $\int\varphi(x)\delta_{1D}(ax)dx = \frac{\varphi(0)}{|a|}$. By partitioning the above integral over $\mathbf{y}\in\mathbb{R}^3$ into integrals over spheres centered at $\mathbf{x}^{\text{ant}}$, and using the definition of the one dimensional delta distribution, one obtains that

$$E_{\text{Born}}(\mathbf{x}^{\text{ant}};\mathbf{x}^{\text{ant}},t) = w_b'' *_t \frac{(\mathbf{R}_{3\text{D}}\,u^{3D})(\mathbf{x}^{\text{ant}},c_0t/2)}{32\pi^2c_0^3(t/2)^2} \tag{13}$$

with

$$(\mathbf{R}_{3\text{D}}\,u^{3D})(\mathbf{x}^{\text{ant}},r) := \int_{|\mathbf{x}^{\text{ant}}-\mathbf{y}|=r} u^{3D}(\mathbf{y})\;dS(\mathbf{y}) \tag{14}$$

denoting the (three dimensional) spherical Radon transform. This is the basic equation of GPR, that relates the unknown function u^{3D} with the scattered data measured in zero offset mode.

The Radiating Reflectors Model

In our application (see Figure 1) the distances between the antenna position $\mathbf{x}^{\text{ant}}$ and the positions $\mathbf{y}$ of the reflectors are relatively large. In this case, multiplication by t and convolution with w_b'' in (13) can be (approximately) interchanged, that is, we have the approximation

$$(8\pi c_0^2 t)E_{\text{Born}}(\mathbf{x}^{\text{ant}};\mathbf{x}^{\text{ant}},2t) \simeq \Phi(\mathbf{x}^{\text{ant}},t) =: w_b'' *_t \frac{(\mathbf{R}_{3\text{D}}\,u^{3D})(\mathbf{x}^{\text{ant}},c_0t)}{4\pi c_0 t}\,. \tag{15}$$

One notes that Φ is the solution at position $\mathbf{x}^{\text{ant}}$ of the wave equation

$$\left(\frac{1}{c_0^2}\partial_t^2-\Delta\right)\Phi(\mathbf{x},t) = w_b''(t)u^{3D}(\mathbf{x})\,, \qquad (\mathbf{x},t)\in\mathbb{R}^3\times\mathbb{R}\,. \tag{16}$$

Equation (16) is named the *radiating (or exploding) reflectors model*, as the inhomogeneity u^{3D} now appears as active source in the wave equation.

Formulation of the Inverse Problem

Equation (15) relates the unknown function $u^{3D}(\mathbf{x})$ with the data $\Phi(\mathbf{x}^{\text{ant}}, t)$. Due to the convolution with the function w_b'', which does not contain high frequency components (see Figure 2), the exact reconstruction of u^{3D} is hardly possible. It is therefore common to apply migration, which is designed to invert the spherical Radon transform.

When applying migration to the data defined in (15), one reconstructs a band-limited approximation of u^{3D}. Indeed, from [28, Proposition 2.2], it follows that

$$\Phi(\mathbf{x}^{\text{ant}}, t) = \frac{(\mathbf{R}_{\text{3D}}\, u_b^{3D})(\mathbf{x}^{\text{ant}}, c_0 t)}{t}\,, \tag{17}$$

where

$$u_b^{3D}(\mathbf{x}) := -\frac{\pi}{8\pi c_0} \int_{\mathbb{R}^3} \frac{w_b'''(|\mathbf{y}|)}{|\mathbf{y}|}\, u^{3D}(\mathbf{x}-\mathbf{y})d\mathbf{y}\,, \qquad \mathbf{x} \in \mathbb{R}^3\,. \tag{18}$$

Therefore, the data $t\Phi(\mathbf{x}^{\text{ant}}, t)$ can be viewed a the spherical Radon transform of the band-limited reflectivity function $u_b^{3D}(\mathbf{x})$, and application of migration to the data $t\Phi(\mathbf{x}^{\text{ant}}, t)$ will reconstruct the function $u_b^{3D}(\mathbf{x})$.

A characteristic of our application (see Figure 1) is that the radar antenna is moved along a one dimensional path, that is, only the two dimensional data set

$$v(x^{\text{ant}}, t) := t\Phi\big((x^{\text{ant}}, 0, 0), t\big)\,, \qquad \text{with } (x^{\text{ant}}, t) \in \mathbb{R} \times (0, \infty)\,,$$

is available from which one can recover at most a function with two degrees of freedom. Therefore, we make the assumption that the support of the function u_b^{2D} is approximately located in the plane $\{(x_1, x_2, x_3) : x_3 = 0\}$, that is, we assume

$$u_b^{3D}(x_1, x_2, x_3) = u_b^{2D}(x_1, x_2)\ \delta_{\text{1D}}(x_3)\,, \qquad \text{with} \qquad \mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^2 \times \mathbb{R}\,.$$

Together with (17) this leads to the equation

$$v(x^{\text{ant}}, t) = \big(\mathbf{R}_{\text{2D}}\, u_b^{2D}\big)(x^{\text{ant}}, c_0 t)\,, \qquad (x^{\text{ant}}, t) \in \mathbb{R} \times (0, \infty)\,, \tag{19}$$

where

$$(\mathbf{R}_{\text{2D}}\, u)(x^{\text{ant}}, r) := \int_{|(x^{\text{ant}}, 0) - \mathbf{y}| = r} u\big(\mathbf{y}\big)\ dS(\mathbf{y})\,, \qquad (x^{\text{ant}}, r) \in \mathbb{R} \times (0, \infty)\,, \tag{20}$$

denotes the circular Radon transform (the spherical Radon in two dimensions). Equation (19) is the final equation that will be used to reconstruct the bandlimited reflectivity function $u_b^{2D}(x_1, x_2)$ from data $v(x^{\text{ant}}, r)$.

5.2 Migration versus Nonlinear Focusing

If the values $(\mathbf{R}_{2\mathrm{D}}\, u_b^{2D})(x^{\mathrm{ant}}, r)$ in (19) were known for all $x^{\mathrm{ant}} \in \mathbb{R}$ and all $r > 0$, then u_b^{2D} could be reconstructed by means of explicit reconstruction formulas. At least two types of theoretically exact formulas for recovering u_b^{2D} have been derived: Temporal back-projection and Fourier domain formulas [1, 18, 37, 44]. These formulas and their variations are known as migration, backprojection, or synthetic focusing techniques.

The Limited Data Problem

In practise it is not appropriate to assume $(\mathbf{R}_{2\mathrm{D}}\, u_b^{2D})(x^{\mathrm{ant}}, t)$ is known for all $x^{\mathrm{ant}} \in \mathbb{R}$, and the antenna positions and acquisition times have to be restricted to domains $(-X, X)$ and $(0, R/c_0)$, respectively. We model the available partial data by

$$v_{\mathrm{cut}}(x^{\mathrm{ant}}, r) := w_{\mathrm{cut}}(x^{\mathrm{ant}}, r)\, (\mathbf{R}_{2\mathrm{D}}\, u_b^{2D})(x^{\mathrm{ant}}, r)\,, \quad \text{with } (x^{\mathrm{ant}}, r) \in (-X, X) \times (0, R)\,, \tag{21}$$

where w_{cut} is a smooth cutoff function that vanishes outside the domain $(-X, X) \times (0, R)$. Without a–priori knowledge, the reflectivity function u_b^{2D} cannot be exactly reconstructed from the incomplete data (21) in a stable way (see [35]). It is therefore common to apply migration techniques just to the partial data and to consider the resulting image as approximate reconstruction.

Applying Kirchhoff migration to the partial data (21) leads to

$$u_{\mathrm{Km}}(x_1, x_2) := (\mathbf{R}_{2\mathrm{D}}^{*}\, v_{\mathrm{cut}})(x_1, x_2) := \int_{-X}^{X} v_{\mathrm{cut}} \left(x^{\mathrm{ant}}, \sqrt{(x^{\mathrm{ant}} - x_1)^2 + x_2^2} \right) dx^{\mathrm{ant}}\,.$$

With Kirchhoff migration, the horizontal resolution at location $(0, d)$ is given by $c_0 d/(2Xb)$ (see [4, Appendix A.1] for a derivation).

Incorporating a–priori knowledge via non-linear inversion, however, may be able to increase the resolution. Below we will demonstrate that this is indeed the case for sparsity regularization using a Haar wavelet basis. A heuristic reason is that sparse objects (reconstructed with sparse regularization) tend to be less blurred than images reconstructed by linear methods.

Application of Sparsity Regularization

For the sake of simplicity we will only consider Tikhonov regularization with $\mathcal{R}_1$ penalty term and uniform weights, leading to the regularization functional

$$\mathcal{T}_{\alpha, v^\delta}(u) := \|\mathbf{R}_{2\mathrm{D}}\, u - v^\delta\|^2 + \alpha \sum_{\lambda \in \Lambda} |\langle \phi_\lambda, u \rangle|\,, \tag{22}$$

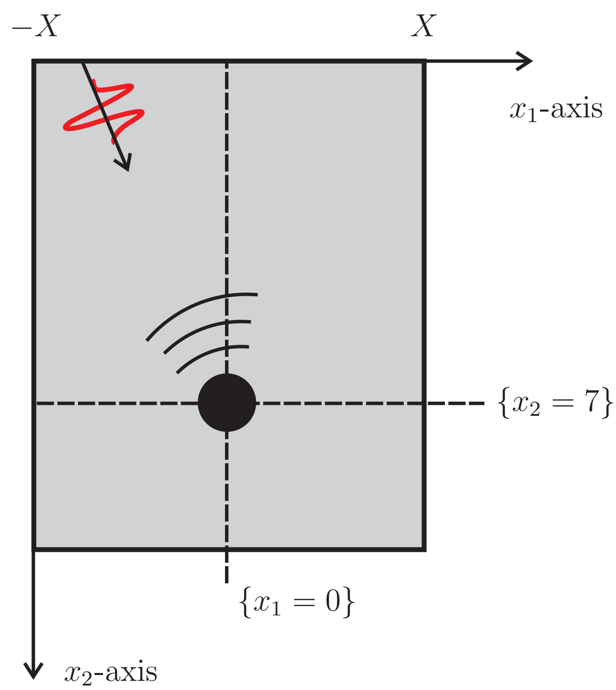

Figure 3: **Geometry in the numerical experiment.** Data $v(x^{\mathrm{ant}}, r)$, caused by a small scatterer positioned at location $(0, 7\mathrm{m})$, are simulated for $(x^{\mathrm{ant}}, r) \in (-X, X) \times (0, R)$ with $X = 2\mathrm{m}$ and $R = 12\mathrm{m}$.

where $(\phi_\lambda)_{\lambda\in\Lambda}$ is a Haar wavelet basis and α is the regularization parameter. Here u and v^δ are elements of the Hilbert spaces

$$\begin{aligned} U &:= \left\{u \in L^2\big(\mathbb{R}^2\big) : \operatorname{supp}(u) \subset \overline{(-X, X) \times (0, R)}\right\}, \\ V &:= L^2\big((-X, X) \times (0, R)\big). \end{aligned}$$

The circular Radon transform $\mathbf{R}_{2\mathrm{D}}$, considered as operator between U and V, is easily shown to be bounded linear (see, e.g., [43, Lemma 3.79].)

For the minimization of (22), we apply the iterative thresholding algorithm (7), which in our context reads as

$$u_{n+1} := \mathbf{S}_{\mu\alpha,1}\big(u_n + \mu\,\mathbf{R}_{2\mathrm{D}}^*(v^\delta - \mathbf{R}_{2\mathrm{D}}\,u_n)\big)\,. \tag{23}$$

Here $\mathbf{S}_{\mu\alpha,1}$ is the shrinkage operator defined by (6) and (5), and μ is a positive parameter such that $\mu\|\mathbf{R}_{2\mathrm{D}}^*\,\mathbf{R}_{2\mathrm{D}}\| < 1$.

5.3 Numerical Examples

In our numerical examples we choose $X = 2\mathrm{m}$ and $R = 12\mathrm{m}$. The scatterer u is the characteristic function of a small disc located at position $(0, d)$ with

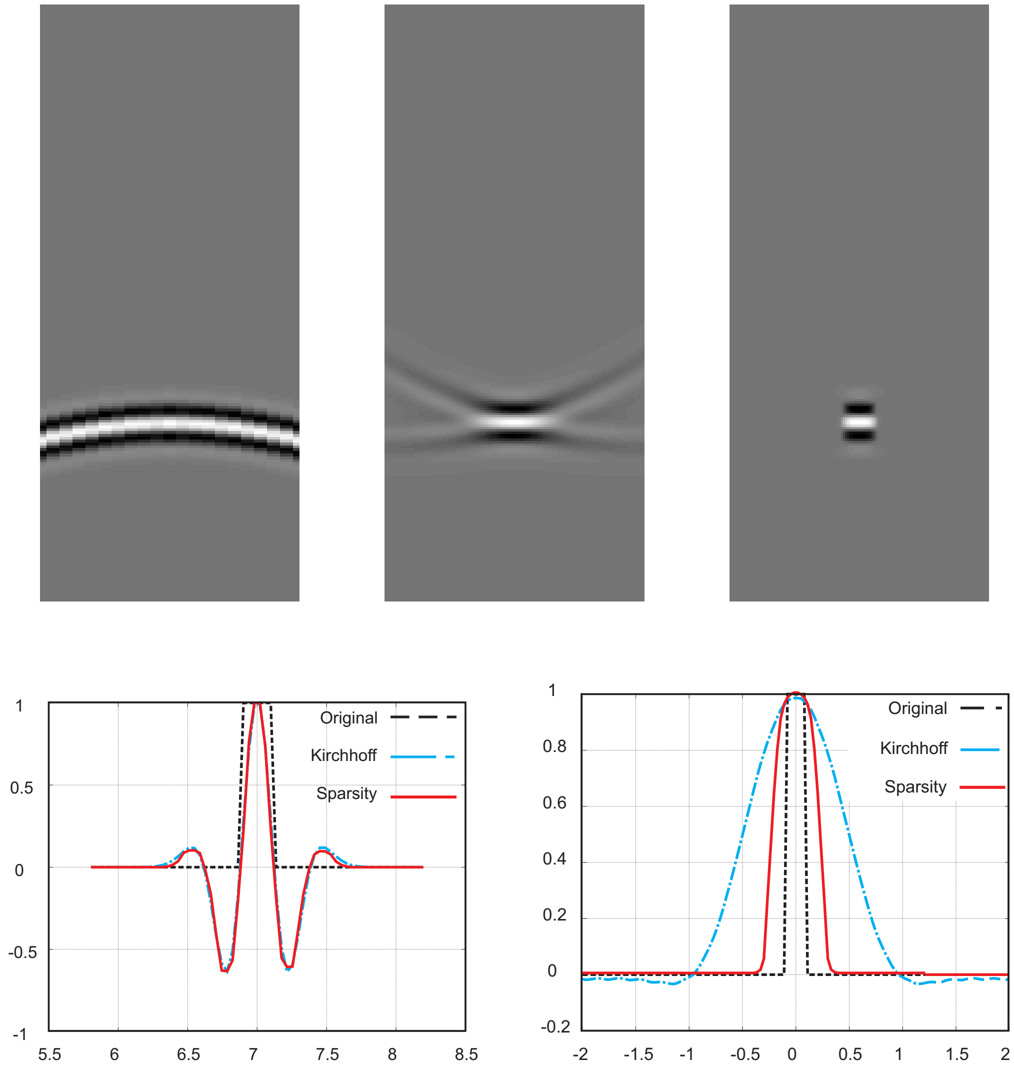

Figure 4: **Exact data experiment.** Top left: Data. Top middle: Reconstruction by Kirchhoff migration. Top right: Reconstruction with sparsity regularization. Bottom: Vertical and horizontal profiles of the reconstructions.

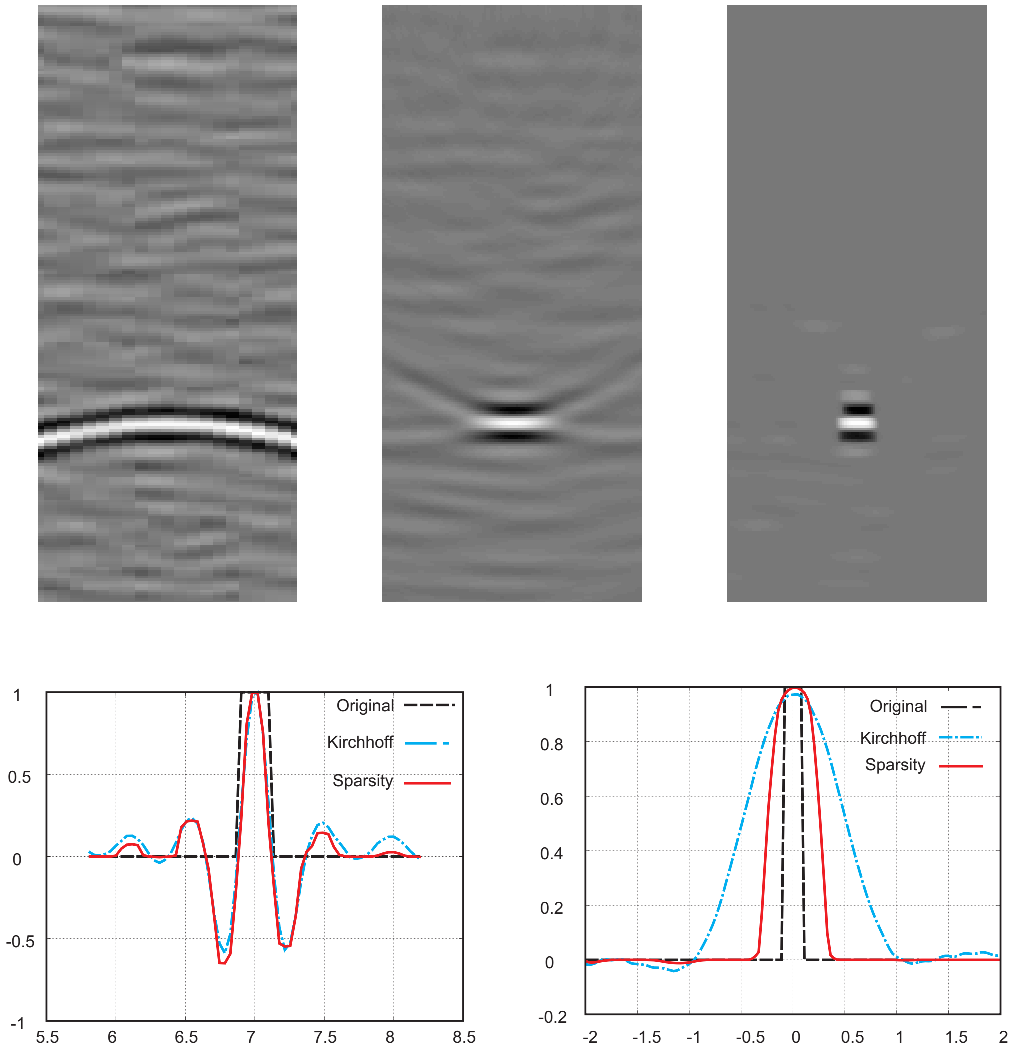

Figure 5: **Noisy data experiment.** Top left: Data. Top middle: Reconstruction by Kirchhoff migration. Top right: Reconstruction with sparsity regularization. Bottom: Vertical and horizontal profiles of the reconstructions.

$d = 7$, see Figure 3. We assume that the emitted Radar signal is a Ricker wavelet w_b with a central frequency of 250MHz (compare with Figure 2). The data $v(x^{\mathrm{ant}}, r)$ are generated by numerically convolving $\mathbf{R}_{2\mathrm{D}}\, u$ with the second derivative of the Ricker wavelet.

The reconstructions obtained with Kirchhoff migration and with sparsity regularization are depicted in Figure 4. Both methods show good resolution in the vertical direction (often called axial or range resolution). The horizontal resolution (lateral or cross-range resolution) of the scatterer, however, is significantly improved by sparsity regularization. This shows that sparsity regularization is indeed able to surpass the resolution limit $c_0 d/(2Xb)$ of linear reconstruction techniques.

In order to demonstrate the stability with respect to data perturbations, we also perform reconstructions after adding Gaussian noise and clutter. Clutter occurs from multiple reflections on fixed structures and reflections resulting from the inhomogeneous background [12]. A characteristic property of clutter is that is has similar spectral characteristics as the emitted radar signal

The reconstruction results from data with clutter and noise added are depicted in Figure 5. Again, sparsity regularization shows better horizontal resolution than Kirchhoff migration. Moreover, the image reconstructed with sparsity regularization is less noisy.

5.4 Application to Real Data

Radar measurements were performed with a 400MHz antenna (RIS One GPR instrument). The investigated area was a complex avalanche deposit near Salzburg, Austria. The recorded data are shown in Figure 6. In the numerical reconstruction we choose an aperture of $X = 3.3$m and a time window of $R/c_0 = 50$ns. The extracted data are depicted in the left image in Figure 7. One clearly sees a diffraction hyperbola stemming from a scatterer in the subsurface. Moreover, the data agree very well with the simulated data depicted in the left image in Figure 5.

The reconstruction results with Kirchhoff migration and with sparsity regularization are depicted in Figure 7. The regularization parameter α is chosen as 0.02, and the scaling parameter μ is chosen in such a way, that $\mu \|\mathbf{R}_{2\mathrm{D}}^* \mathbf{R}_{2\mathrm{D}}\|$ is only slightly smaller than 1.

Acknowledgement

This work has been supported by the Austrian Science Fund (FWF) within the national research networks Industrial Geometry, project 9203-N12, and Photoacoustic Imaging in Biology and Medicine, project S10505-N20. The

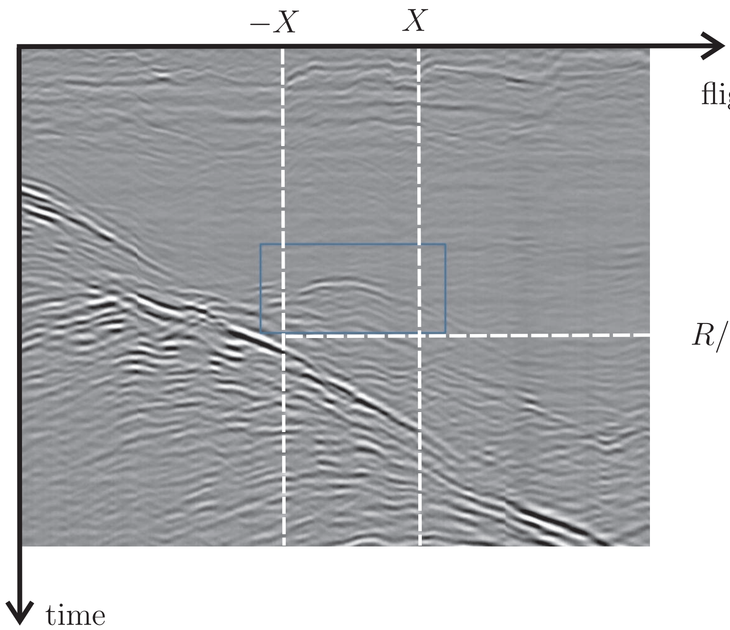

Figure 6: **Measured radar data.** For the numerical reconstruction only the partial data $\Phi((x^{\text{ant}},0,0),t)$, with $(x^{\text{ant}},t) \in (-X,X) \times (0,R/c_0)$ where $X = 3.3$m and $R/c_0 = 50$ns, have been used.

authors thank Sylvia Leimgruber (alpS - Center for Natural Hazard Management in Innsbruck) and Harald Grossauer (University Innsbruck) for providing providing real life data sets.

References

[1] L. E. Andersson. On the determination of a function from spherical averages. *SIAM J. Math. Anal.*, 19(1):214–232, 1988.

[2] N. Bleistein, J. K. Cohen, and J. W. Stockwell Jr. *Mathematics of multidimensional seismic imaging, migration, and inversion*, volume 13 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, New York, 2001. Geophysics and Planetary Sciences.

[3] T. Bonesky. Morozov's discrepancy principle and Tikhonov-type functionals. *Inverse Probl.*, 25(1):015015, 11, 2009.

[4] L. Borcea, G. Papanicolaou, and C. Tsogka. Interferometric array imaging in clutter. *Inverse Probl.*, 21(4):1419–1460, 2005.

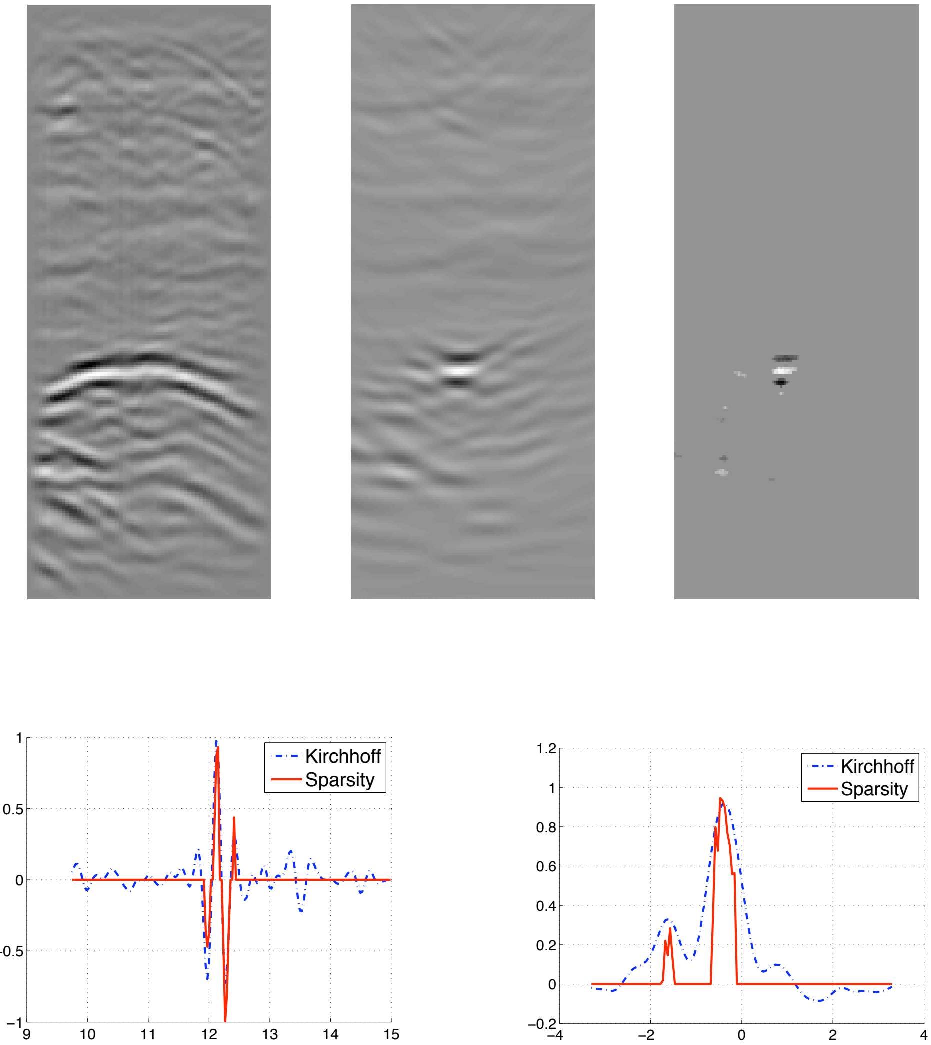

Figure 7: **Reconstruction from real data.** Top left: Data. Top middle: Reconstruction by Kirchhoff migration. Top right: Reconstruction with sparsity regularization. Bottom: Vertical and horizontal profiles of the reconstructions.

[5] K. Bredies and D. Lorenz. Minimization of Non-smooth, Non-convex Functionals by Iterative Thresholding. DFG-Schwerpunktprogramm 1324, Preprint 10, 2009.

[6] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Probl.*, 20(5):1411–1421, 2004.

[7] E. J. Candès and J. Romberg. ℓ_1-MAGIC: Recovery of Sparse Signals via Convex Programming. Technical report, 2005. Available at `http://www.acm.caltech.edu/l1magic`.

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

[9] J. Claerbout and F. Muir. Robust modeling of erratic data. *Geophysics*, 38:826–844, 1973.

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

[11] R. Courant and D. Hilbert. *Methods of Mathematical Physics*, volume 2. Wiley-Interscience, New York, 1962.

[12] D. Daniels. *Ground Penetrating Radar*. The Institution of Electrical Engineers, London, 2004.

[13] D. Daniels. Ground Penetrating Radar. In M. Skolnick, editor, *Radar Handbook*, chapter 21, pages 21.1–21.41. The McGraw-Hill Companies, 2008.

[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(11):1413–1457, 2004.

[15] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization. *Proc. Natl. Acad. Sci. USA*, 100(5):2197–2202 (electronic), 2003.

[16] I. Ekeland and R. Temam. *Analyse convexe et problèmes variationnels.* Dunod, 1974. Collection Études Mathématiques.

[17] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[18] J. A. Fawcett. Inversion of n-dimensional spherical averages. *SIAM J. Appl. Math.*, 45(2):336–341, 1985.

[19] D. Finch and Rakesh. The spherical mean value operator with centers on a sphere. *Inverse Probl.*, 23(6):37–49, 2007.

[20] F. Frühauf, A. Heilig, M. Schneebeli, W. Fellin, and O. Scherzer. Experiments and algorithms to detect snow avalanche victims using airborne ground-penetrating radar. *IEEE Trans. Geosci. Remote Sens.*, 47(7), 2009.

[21] M. Grasmair. Non-convex sparse regularisation. *J. Math. Anal. Appl.*, 2009. To appear.

[22] M. Grasmair. Well-posedness and convergence rates for sparse regularization with sublinear l^q penalty term. *Inverse Probl. Imaging*, 3(3):383–387, 2009.

[23] M. Grasmair, M. Haltmeier, and O. Scherzer. Sparse regularization with l^q penalty term. *Inverse Probl.*, 24(5):055020 (13pp), 2008.

[24] M. Grasmair, M. Haltmeier, and O. Scherzer. Necessary and sufficient conditions for linear convergence of ℓ^1-regularization. Reports of FSP S105 - "Photoacoustic Imaging" 18, University of Innsbruck, Austria, August 2009. Submitted.

[25] M. Grasmair, M. Haltmeier, and O. Scherzer. The residual method for regularizing ill-posed problems. Reports of FSP S105 - "Photoacoustic Imaging" 14, University of Innsbruck, Austria, May 2009. Submitted.

[26] C. W. Groetsch. *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind.* Pitman, Boston, 1984.

[27] M. Haltmeier, R. Kowar, and O. Scherzer. Computer aided location of avalanche victims with ground penetrating radar mounted on a helicopter. In *[32]*, pages 19–28, 2005.

[28] M. Haltmeier, O. Scherzer, and G. Zangerl. Influence of detector bandwidth and detector size to the resolution of photoacoustic tomagraphy. In F. Breitenecker and I. Troch, editors, *Argesim Report no. 35: Proceedings Mathmod 09 Vienna*, pages 1736–1744, 2009.

[29] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer. A convergence rates result in Banach spaces with non-smooth operators. *Inverse Probl.*, 23(3):987–1010, 2007.

[30] V. K. Ivanov, V. V. Vasin, and V. P. Tanana. *Theory of linear ill-posed problems and its applications*. Inverse and Ill-posed Problems Series. VSP, Utrecht, second edition, 2002. Translated and revised from the 1978 Russian original.

[31] P. Kuchment and L. A. Kunyansky. Mathematics of thermoacoustic and photoacoustic tomography. *European J. Appl. Math.*, 19:191–224, 2008.

[32] F. Lenzen, O. Scherzer, and M. Vincze, editors. *Digital Imaging and Pattern Recognition*, Proceedings of 30th Workshop of the Austrian Association for Pattern Recognition, 2005. Can be ordered at `http://www.ocg.at/bookshop/`.

[33] S. Levy and T. Fullagar. Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution. *Geophysics*, 46:1235–1243, 1981.

[34] D. Lorenz. Convergence rates and source conditions for Tikhonov regularization with sparsity constraints. *J. Inverse Ill-Posed Probl.*, 16(5):463–478, 2008.

[35] A. K. Louis and E. T. Quinto. Local tomographic methods in sonar. In *Surveys on solution methods for inverse problems*, pages 147–154. Springer, Vienna, 2000.

[36] A. Neubauer. On converse and saturation results for Tikhonov regularization of linear ill-posed problems. *SIAM J. Numer. Anal.*, 34:517–527, 1997.

[37] S. J. Norton and M. Linzer. Ultrasonic reflectivity imaging in three dimensions: Exact inverse scattering solutions for plane, cylindrical and spherical apertures. *IEEE Trans. Biomed. Eng.*, 28(2):202–220, 1981.

[38] D. Oldenburg, T. Scheuer, and S. Levy. Recovery of the acoustic impedance from reflection seismograms. *Geophysics*, 48:1318–1337, 1983.

[39] S. K. Patch and O. Scherzer. Special section on photo- and thermoacoustic imaging. *Inverse Probl.*, 23:S1–S122, 2007.

[40] J. Renegar. *A mathematical view of interior-point methods in convex optimization*. MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001.

[41] E. Resmerita. Regularization of ill-posed problems in Banach spaces: convergence rates. *Inverse Probl.*, 21(4):1303–1314, 2005.

[42] F. Santosa and W. W. Symes. Linear inversion of band-limited reflection seismograms. *SIAM J. Sci. Comput.*, 7(4):1307–1330, 1986.

[43] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, New York, 2009.

[44] R. H. Stolt. Migration by Fourier transform. *Geophysics*, 43:23–48, 1978.

[45] W. W. Symes. The seismic reflection inverse problem. *Inverse Probl.*, 2009. To appear.

[46] C. A. Zarzer. On Tikhonov regularization with non-convex sparsity constraints. *Inverse Probl.*, 25:025006, 2009.

Index