National Research Network S92

Industrial Geometry

http://www.industrial-geometry.at

NRN Report No. 86

Non-convex Sparse Regularization

Markus Grasmair

May 2009

Der Wissenschaftsfonds.

Non-convex Sparse Regularisation

Markus Grasmair

Department of Mathematics
University of Innsbruck
Technikerstr. 21a
6020 Innsbruck, Austria

May 29, 2009

Abstract

We study the regularising properties of Tikhonov regularisation on the sequence space ℓ^2 with weighted, non-quadratic penalty term acting separately on the coefficients of a given sequence. We derive sufficient conditions for the penalty term that guarantee the well-posedness of the method, and investigate to which extent the same conditions are also necessary. A particular interest of this paper is the application to the solution of operator equations with sparsity constraints. Assuming a linear growth of the penalty term at zero, we prove the sparsity of all regularised solutions. Moreover, we derive a linear convergence rate under the assumptions of even faster growth at zero and a certain injectivity of the operator to be inverted. These results in particular cover non-convex ℓ^p regularisation with $0 < p < 1$.

MSC: 65J20,47A52;

Keywords: Tikhonov regularisation, sparsity, convergence rates.

1 Introduction

Regularisation with sparsity constraints is an impressingly effective method for the solution of operator equations

$$Ax = y\,,$$

when it is known that the solution only contains a small number of significant coefficients. The idea is that, instead of minimising the classical Tikhonov functional $\mathcal{T}_\alpha(x, y) = \|Ax-y\|^2 + \alpha\|x\|_2^2$, one increases the penalisation of small coefficients of x while at the same time decreasing the penalisation of the large ones. Following [7], this can be achieved by replacing the ℓ^2 term used for the regularisation by an ℓ^p norm with $p < 2$. The corresponding regularisation functional then reads as

$$\mathcal{T}_\alpha(x, y) = \|Ax - y\|^2 + \alpha\|x\|_p^p \qquad \text{with } p < 2\,.$$

Applications and solution algorithms for such problems can be found in [4, 5, 6, 7, 18]. The regularising properties of this type of functionals have been analysed in [7, 10, 11, 14, 17, 20]. In addition, the related constrained optimisation problem $\|Ax - y\|^2 \to \min$ subject to $\|x\|_p \le \delta$ has been studied in the context of compressed sensing [3, 8]. Moreover, we refer to [2], where an overview of sparse regularisation is given.

In this paper we study more general, weighted regularisation functionals of the form

$$\mathcal{T}_\alpha(x, y) = \|Ax - y\|^2 + \alpha \sum_\lambda w_\lambda \phi(x_\lambda)$$

with $\phi\colon \mathbb{R} \to [0, +\infty]$ and $w_\lambda > 0$. We derive necessary and sufficient conditions for $\mathcal{T}_\alpha$ to define a well-posed regularisation method. The main condition turns out to be the behaviour of ϕ at zero. Quadratic or faster growth implies the well-posedness of the method—though slower growth is also possible if it is compensated by the weights. Linear growth of the function ϕ at zero implies that the minimisers of $\mathcal{T}_\alpha$ are necessarily sparse. Finally, we derive a linear convergence rate in the case of sublinear growth under the additional assumption that the operator A satisfies a kind of finite basis injectivity property.

2 Overview of the Results

Let $\ell^2 = \ell^2(\Lambda)$ for some countable index set Λ, and let Y be some Hilbert space. We study the stable solution of the equation $Ax = y$ by means of Tikhonov regularisation, where $A\colon \ell^2 \to Y$ is a bounded linear operator. For $\alpha > 0$ we consider the functional $\mathcal{T}_\alpha\colon \ell^2 \times Y \to [0, +\infty]$,

$$\mathcal{T}_\alpha(x, y) := \|Ax - y\|_Y^2 + \alpha \mathcal{R}(x)\,,$$

where the regularisation term $\mathcal{R}\colon \ell^2 \to [0, +\infty]$ has the form

$$\mathcal{R}(x) = \sum_\lambda w_\lambda \phi(x_\lambda)\ . \tag{1}$$

Here $\phi\colon \mathbb{R} \to [0, +\infty]$ is some non-negative function and the weights w_λ satisfy $w_\lambda > 0$ for every $\lambda \in \Lambda$.

The first task is, to formulate conditions on ϕ and the weights w_λ that guarantee that the functional $\mathcal{T}_\alpha(\cdot, y)$ admits a minimiser for every $\alpha > 0$ and $y \in Y$. This is the case, if the functional $\mathcal{R}$ is weakly lower semi-continuous and weakly coercive; the latter condition means that $\|x\|_{\ell^2} \to \infty$ implies $\mathcal{R}(x) \to \infty$. In this paper we prove weak lower semi-continuity and weak coercivity of $\mathcal{R}$ under the following conditions $C1$–$C3$ (see Propositions 3.1 and 3.4):

$C1$ The mapping $\phi\colon \mathbb{R} \to [0, +\infty]$ is lower semi-continuous and $\phi(0) = 0$.

$C2$ We have $\lim_{|t|\to\infty} \phi(t) = +\infty$.

$C3$ There exist $p \geq 1$ and $q \in (0, +\infty]$ satisfying $p - 1 = 1/q$ such that $(w_\lambda^{-1})_{\lambda \in \Lambda} \in \ell^q$ and, for some $C > 0$,

$$\phi(t) \geq \frac{C|t|^{2p}}{1 + |t|^{2p}} \qquad \text{for every } t \in \mathbb{R} \,. \tag{2}$$

In addition to the sufficiency of these conditions, we investigate to which extent they are necessary for the weak lower semi-continuity and the weak coercivity of $\mathcal{R}$. We prove the necessity of conditions $C1$ and $C2$ and derive some necessary properties of the weights w_λ and the function ϕ (see Propositions 3.1 and 3.6).

Moreover we consider the special case where the weights w_λ are bounded above. We show that in this situation (2) has to be satisfied with $p = 1$, and thus we obtain a complete characterisation of weakly lower semi-continuous and weakly coercive functionals of the form (1) with bounded weights. This generalises and completes the results recently derived in [1], where only the case of constant weights has been investigated.

Also in [1], the question has been asked, whether the functional $\mathcal{R}$ satisfies the *Radon–Riesz property*, also known as *Kadec'–Klee property* (see [15]). This property requires that every sequence $(x^{(k)})_{k \in \mathbb{N}} \subset \ell^2$, which converges weakly to some $x \in \ell^2$ in such a way that $\mathcal{R}(x^{(k)}) \to \mathcal{R}(x) < \infty$, already satisfies $\|x^{(k)} - x\|_{\ell^2} \to 0$. This is important for the derivation of convergence and stability theorems for Tikhonov regularisation, as it allows one to infer results in the norm topology instead of merely the weak topology. Generalising [1], we prove in Proposition 3.7 that the Radon–Riesz property is already a consequence of conditions $C1$–$C3$ and thus naturally satisfied.

As a consequence of the considerations above, it follows that, under conditions $C1$–$C3$, the proposed functional $\mathcal{T}_\alpha$ satisfies the main properties of a regularisation method. The weak lower semi-continuity and weak coercivity of $\mathcal{R}$ imply the existence of minimisers for every $y \in Y$ and $\alpha > 0$ (see Proposition 4.1). The Radon–Riesz property implies stability of the method under perturbations of y and α (see Proposition 4.2). Also, it implies the convergence of minimisers x_α^δ of $\mathcal{T}_\alpha(\cdot, y^\delta)$ to solutions of $Ax = y$ provided the noise level $\delta = \|y^\delta - y\|$ and the regularisation parameter α converge to zero in a suitable manner (see Proposition 4.3). These results provide further generalisations of similar statements that have first been derived for weighted ℓ^p regularisation with $p \geq 1$ in [14, 17], for ℓ^p regularisation with $0 < p < 1$ and constant weights in [10, 20], and for general symmetric ϕ but constant weights in [1].

In order to enforce the sparsity of the regularised solutions, it is necessary to introduce a stronger growth condition for ϕ at zero. This condition $C3'$ below replaces the quadratic or slower growth of ϕ required in condition $C3$ by at least linear growth. In Proposition 4.5 we prove that this condition implies the sparsity of every minimiser of the functional $\mathcal{T}_\alpha$.

$C3'$ We have $\inf_\lambda w_\lambda > 0$ and there exists $C > 0$ such that

$$\phi(t) \geq \frac{C|t|}{1+|t|} \qquad \text{for every } t \in \mathbb{R}\,.$$

For the derivation of linear convergence rates we propose an even stronger growth condition at zero and a weak regularity condition for the function ϕ. To that end recall that the *lower Dini derivatives* of a function $\rho\colon \mathbb{R} \to [0, +\infty]$ at $t \in \mathbb{R}$ are defined as (see [13, Def. 17.2])

$$D_+\rho(t) = \liminf_{h\to 0^+} \frac{\rho(t+h)-\rho(t)}{h}\,, \qquad D_-\rho(t) = \liminf_{h\to 0^-} \frac{\rho(t+h)-\rho(t)}{h}\,.$$

$C4$ For every $t \in \mathbb{R}$ with $\phi(t) < +\infty$ we have

$$D_+\phi(t) > -\infty\,, \qquad \text{and} \qquad D_-\phi(t) < +\infty\,.$$

Moreover

$$D_+\phi(0) = +\infty\,, \qquad \text{and} \qquad D_-\phi(0) = -\infty\,.$$

Following the argumentation in [12], where constrained ℓ^p regularisation with $0 < p < 1$ has been considered, we add to $C1$–$C4$ the condition that the equation $Ax = y$ has a unique $\mathcal{R}$-minimising solution $x^\dagger$, which is sparse, that is, the support $\Omega := \operatorname{supp}(x^\dagger) := \{\lambda \in \Lambda : x^\dagger_\lambda \neq 0\}$ is finite. In addition, we assume that the restriction of the operator A to $\ell^2(\Omega)$ is injective. This is a special instance of the finite basis injectivity property proposed in [14]. We prove that these conditions imply the linear convergence of minimisers x_α^δ of $\mathcal{T}_\alpha(\cdot, y^\delta)$ to $x^\dagger$ as $\alpha \sim \delta = \|y - y^\delta\| \to 0$ (see Theorem 5.1).

Linear convergence rates for non-convex regularisation have already been derived in [1, 10], albeit with the much stronger range condition $e_\lambda \in \operatorname{Range} A^*$ for every $\lambda \in \Omega$ with $(e_\lambda)_{\lambda\in\Lambda}$ denoting the set of standard basis vector of ℓ^2. At the same time, a rate of order $O(\sqrt{\delta})$ has been proven in [20] for ℓ^p regularisation with $0 < p < 1$. There, the less restrictive condition has been assumed that there exists some $\omega \in Y$ such that $|x^\dagger_\lambda|^{2-p}(A^*\omega)_\lambda = x^\dagger_\lambda$ for every $\lambda \in \Lambda$. It has been noted in [12, 16] that this range condition is a consequence of the injectivity condition required in our convergence rates result.

Now we present some examples of functions ϕ to which our results apply. For simplicity, we always assume that the chosen weights w_λ are uniformly bounded below, that is, $\inf_\lambda w_\lambda > 0$.

Example 2.1 (ℓ^r Regularisation). Here,

$$\phi(t) = |t|^r \qquad \text{for some } r > 0\,.$$

The mapping ϕ is lower semi-continuous, $\phi(0) = 0$, and $\lim_{|t|\to\infty} \phi(t) = +\infty$, proving $C1$ and $C2$. Condition $C3$ is satisfied, if $r \leq 2$; for $r > 2$ we require in addition that $(w_\lambda^{-1})_{\lambda\in\Lambda} \in \ell^q$ with $1/q = r/2 - 1$. If $r \leq 1$, then condition $C3'$ holds. Finally, condition $C4$ is satisfied for $r < 1$. ■

Example 2.2. Assume that

$$\phi(t) = \log(|t| + 1)\ .$$

Then conditions $C1$–$C3$ and $C3'$ are satisfied, while $C4$ does not hold. ∎

Example 2.3 (Positivity Constraints). For any $\phi\colon \mathbb{R} \to [0, +\infty]$ define $\phi_+ : \mathbb{R} \to [0, +\infty]$ by

$$\phi_+(t) := \begin{cases} \phi(t)\,, & \text{if } t \geq 0\,, \\ +\infty\,, & \text{if } t < 0\ . \end{cases}$$

Regularisation with ϕ_+ therefore forces the minimisers to stay non-negative. If ϕ satisfies any of the conditions $C1$–$C4$ and $C3'$, then ϕ_+ satisfies the same conditions. ∎

Example 2.4 (Hard Constraints). For any $\phi\colon \mathbb{R} \to [0, +\infty]$ and $b \geq 0$ define $\phi_b\colon \mathbb{R} \to [0, +\infty]$ by

$$\phi_b(t) := \begin{cases} \phi(t)\,, & \text{if } |t| \leq b\,, \\ +\infty\,, & \text{if } |t| > b\ . \end{cases}$$

This forces the minimisers x of the regularisation functional $\mathcal{T}_\alpha$ to obey the bound $\|x\|_\infty \leq b$. If ϕ satisfies any of the conditions $C1$, $C3$, $C3'$, and $C4$, then ϕ_b satisfies the same conditions. In addition, ϕ_b satisfies condition $C2$. ∎

Example 2.5 (ℓ^0 Regularisation). Define

$$\phi(t) := \begin{cases} 0\,, & \text{if } t = 0\,, \\ 1\,, & \text{if } t \neq 0\ . \end{cases}$$

Then ϕ satisfies the conditions $C1$, $C3$, and $C4$. The condition $C2$, however, is not satisfied, and thus the coercivity of $\mathcal{R}$ does not hold.

On the other hand, if we impose in addition a hard constraint $b > 0$, that is, we replace ϕ by the functional (see Example 2.4)

$$\phi_b(t) := \begin{cases} 0\,, & \text{if } t = 0\,, \\ 1\,, & \text{if } 0 < |t| \leq b\,, \\ +\infty\,, & \text{if } |t| > b\,, \end{cases}$$

then all conditions $C1$–$C4$ are met. ∎

3 Properties of the Regularisation Functional

In the following, we investigate the weak lower semi-continuity and the weak coercivity of the regularisation term defined in (1). First we prove that $C1$–$C3$ are sufficient conditions. Then we turn to the question of their necessity. We show that $C1$ and $C2$ are indeed necessary, while we can only derive condition $C3$ with $p = 1$ in case the weights w_λ are assumed to be bounded. Finally, we prove that the Radon–Riesz property of $\mathcal{R}$ is a direct consequence of the conditions $C1$–$C3$.

Proposition 3.1 (Lower Semi-continuity). *Assume that $\mathcal{R}$ is proper. Then the following are equivalent:*

1. *The mapping ϕ is lower semi-continuous.*
2. *The functional $\mathcal{R}$ is lower semi-continuous.*
3. *The functional $\mathcal{R}$ is weakly lower semi-continuous.*

Proof. First note that the implication $3 \implies 2$ is trivial.

In order to show the implication $2 \implies 1$, choose some $x \in \mathrm{Dom}(\mathcal{R})$. Let $t \in \mathbb{R}$ and $t_k \to t$. Choose some $\mu \in \Lambda$ and define $y^{(k)} \in \ell^2$ by $y_\mu^{(k)} = t_k$ and $y_\lambda^{(k)} = x_\lambda$ for $\lambda \neq \mu$. Define moreover $y \in \ell^2$ by $y_\mu = t$ and $y_\lambda = x_\lambda$ for $\lambda \neq \mu$. Then $y^{(k)} \to y$ and therefore

$$\begin{aligned}\liminf_k w_\mu \phi(t_k) &= \liminf_k \bigl[\mathcal{R}(y^{(k)}) - \mathcal{R}(x) + w_\mu \phi(x_\mu)\bigr] \\ &\geq \mathcal{R}(y) - \mathcal{R}(x) + w_\mu \phi(x_\mu) = \phi(t) \ .\end{aligned}$$

Thus ϕ is lower semi-continuous.

For the implication $1 \implies 3$ note that the lower semi-continuity of the mapping ϕ implies that for every finite set $\Lambda' \subset \Lambda$ the mapping $x \mapsto \sum_{\lambda\in\Lambda'} w_\lambda \phi(x_\lambda)$ is weakly lower semi-continuous. Since ϕ is non-negative, we have

$$\mathcal{R}(x) = \sum_{\lambda\in\Lambda} w_\lambda \phi(x_\lambda) = \sup\Bigl\{ \sum_{\lambda\in\Lambda'} w_\lambda \phi(x_\lambda) : \Lambda' \subset \Lambda \text{ is finite } \Bigr\} \ .$$

Therefore the mapping $\mathcal{R}$ is the supremum of a family of weakly lower semi-continuous functions and therefore itself weakly lower semi-continuous. □

Remark 3.2. The argument for the proof of the implication $2 \implies 1$ is taken from [9, Thm. 6.49], where the same basic idea is applied to the study of lower semi-continuity of integral functionals on Lebesgue spaces. ■

Remark 3.3. In [1], Fatou's Lemma has been used to prove that the conditions 1 and 2 are equivalent to weak *sequential* lower semi-continuity of $\mathcal{R}$. Though yielding a slightly weaker result, this approach has the advantage that it also can be applied when ϕ only satisfies an estimate of the form $\phi(t) \geq -Ct^2$. ■

Proposition 3.4 (Sufficient Conditions for Coercivity). *Assume that the conditions $C2$ and $C3$ are satisfied. Then $\mathcal{R}$ is weakly coercive.*

Proof. Let $K > 0$. Since $(w_\lambda^{-1})_{\lambda\in\Lambda} \in \ell^q$, it follows that $\inf_\lambda w_\lambda > 0$. The condition $\lim_{|t|\to\infty} \phi(t) = +\infty$ therefore implies that there exists some $L > 0$ such that

$$|t| \leq L \qquad \text{whenever } \inf_\lambda w_\lambda \phi(t) \leq K \ .$$

Now let $x \in \ell^2$ satisfy $\mathcal{R}(x) \le K$. Then in particular $|x_\lambda| \le L$ for every $\lambda \in \Lambda$. In case $p = 1$ and $q = +\infty$, we therefore obtain that

$$K \ge \mathcal{R}(x) = \sum_\lambda w_\lambda \phi(x_\lambda) \ge \frac{C \inf_\lambda w_\lambda}{1+L^2} \sum_\lambda x_\lambda^2 = \frac{C \inf_\lambda w_\lambda}{1+L^2} \|x\|_{\ell^2}^2 ,$$

which implies the weak coercivity of $\mathcal{R}$.

In case $p > 1$ and $q < +\infty$, we apply the (reverse) Hölder inequality (see for instance [13, Thm. 13.6]) to obtain the estimate

$$\begin{aligned} K \ge \mathcal{R}(x) = \sum_\lambda w_\lambda \phi(x_\lambda) &\ge \frac{C}{1+L^{2p}} \sum_\lambda w_\lambda |x_\lambda|^{2p} \\ &\ge \frac{C}{1+L^{2p}} \Big(\sum_\lambda w_\lambda^{-q}\Big)^{-1/q} \Big(\sum_\lambda x_\lambda^2\Big)^p . \end{aligned}$$

Thus

$$\|x\|_{\ell^2}^{2p} \le \frac{K(1+L^{2p})}{C} \Big(\sum_\lambda w_\lambda^{-q}\Big)^{1/q} ,$$

which proves the assertion. □

Lemma 3.5. *Let $\rho \colon \mathbb{R} \to [0, +\infty]$ satisfy $\liminf_{t\to 0} \rho(t)/t^2 = 0$. Then there exists a sequence $(x_\lambda)_{\lambda\in\Lambda}$ such that $\sum_\lambda x_\lambda^2 = \infty$ and $\sum_\lambda \rho(x_\lambda) < \infty$.*

Proof. Assume for simplicity of notation that $\Lambda = \mathbb{N}$. Since $\liminf_{t\to 0} \rho(t)/t^2 = 0$, there exists for every $k \in \mathbb{N}$ some $t_k \in \mathbb{R}$ with $0 < |t_k| < 1$ and $\rho(t_k) < 2^{-k} t_k^2$. Choose now an increasing sequence $1 = n_1 < n_2 < \ldots$ such that $1 \le t_k^2(n_{k+1} - n_k) \le 2$ and define $x_\lambda := t_k$ if $n_k \le \lambda < n_{k+1}$. Then

$$\sum_\lambda x_\lambda^2 = \sum_k \sum_{n_k}^{n_{k+1}-1} t_k^2 \ge \sum_k (n_{k+1} - n_k) = +\infty ,$$

while

$$\sum_\lambda \rho(x_\lambda) = \sum_k \sum_{n_k}^{n_{k+1}-1} \rho(t_k) \le \sum_k \sum_{n_k}^{n_{k+1}-1} 2^{-k} t_k^2 \le \sum_k 2^{-k+1} = 2 .$$

□

Proposition 3.6 (Necessary Conditions for Coercivity). *Assume that $\mathcal{R}$ is proper and weakly coercive and that $\phi(\hat{t}) < \infty$ for some $\hat{t} \neq 0$. Then the following hold:*

1. $\inf_\lambda w_\lambda > 0$.
2. $\liminf_{|t|\to 0} \phi(t) = 0$.
3. $\lim_{|t|\to\infty} \phi(t) = +\infty$.

4. *For every $\varepsilon > 0$ we have $\inf_{|t|>\varepsilon} \phi(t) > 0$.*
5. *If $\sup_\lambda w_\lambda < +\infty$, then there exists $C > 0$ such that*

$$\phi(t) \geq \frac{Ct^2}{1+t^2} \qquad \textit{for every } t \in \mathbb{R}\ . \tag{3}$$

Proof. Let $\hat{x} \in \ell^2$ be such that $\mathcal{R}(\hat{x}) < \infty$.

In order to prove Item 1 assume to the contrary that $\inf_\lambda w_\lambda = 0$. Then there exists an infinite subset $\Lambda' \subset \Lambda$ such that $\sum_{\lambda\in\Lambda'} w_\lambda < \infty$. For every finite subset $\Gamma \subset \Lambda$ define now $x^{(\Gamma)} \in \ell^2$ by $x_\lambda^{(\Gamma)} = \hat{t}$ if $\lambda \in \Gamma$ and $x_\lambda^{(\Gamma)} = \hat{x}_\lambda$ if $\lambda \notin \Gamma$. Since $\hat{t} \neq 0$, it follows that $\sup_\Gamma \|x^{(\Gamma)}\|_{\ell^2} = \infty$. On the other hand,

$$\mathcal{R}(x^{(\Gamma)}) = \sum_{\lambda\in\Gamma} w_\lambda \phi(\hat{t}) + \sum_{\lambda\notin\Gamma} w_\lambda \phi(\hat{x}_\lambda) \leq \phi(\hat{t}) \sum_{\lambda\in\Lambda'} w_\lambda + \mathcal{R}(\hat{x})$$

is uniformly bounded, which contradicts the coercivity of $\mathcal{R}$.

Item 2 follows from the fact that $\hat{x} \in \ell^2$ and the estimate

$$\sum_\lambda \phi(\hat{x}_\lambda) \leq \frac{\mathcal{R}(\hat{x})}{\inf_\lambda w_\lambda} < \infty\ .$$

Assume now to the contrary that Item 3 does not hold. Then there exists a sequence t_k with $|t_k| \to \infty$ and $\sup_k \phi(t_k) =: c < \infty$. Now choose some $\mu \in \Lambda$ and define $x^{(k)} \in \ell^2$ by $x_\mu^{(k)} = t_k$ and $x_\lambda^{(k)} = \hat{x}_\lambda$ for $\lambda \neq \mu$. Then $\|x^{(k)}\|_{\ell^2} \to \infty$ while $\sup_k \mathcal{R}(x^{(k)}) \leq \mathcal{R}(\hat{x}) + c < \infty$, which is a contradiction to the coercivity of $\mathcal{R}$.

Now assume that Item 4 does not hold. Then, for some $\varepsilon > 0$, there exists for every $\lambda \in \Lambda$ some $t_\lambda \in \mathbb{R}$ satisfying $|t_\lambda| \geq \varepsilon$ such that $\sum_\lambda w_\lambda \phi(t_\lambda) < \infty$. For every finite subset $\Gamma \subset \Lambda$ we define now $x^{(\Gamma)} \in \ell^2$ by $x_\lambda^{(\Gamma)} = t_\lambda$ if $\lambda \in \Gamma$ and $x_\lambda^{(\Gamma)} = \hat{x}_\lambda$ if $\lambda \notin \Gamma$. Then $\sup_\Gamma \|x^{(\Gamma)}\|_{\ell^2} = \infty$, while $\mathcal{R}(x^{(\Gamma)}) \leq \mathcal{R}(\hat{x}) + \sum_\lambda w_\lambda \phi(t_\lambda)$, again contradicting the coercivity of $\mathcal{R}$.

Now assume that $\sup_\lambda w_\lambda < +\infty$, but (3) does not hold. Since Items 3 and 4 hold, it follows that $\liminf_{|t|\to 0} \phi(t)/t^2 = 0$. From Lemma 3.5 we obtain a sequence $(x_\lambda)_{\lambda\in\Lambda}$ satisfying $\sum_\lambda x_\lambda^2 = +\infty$ and $\sum_\lambda \phi(x_\lambda) =: c < +\infty$. In particular, $\mathcal{R}(x_\lambda) = \sum_\lambda w_\lambda \phi(x_\lambda) \leq c \sup_\lambda w_\lambda$, which, as above, contradicts the coercivity of $\mathcal{R}$. $\square$

Proposition 3.7 (Radon–Riesz Property). *Assume that conditions $C1$–$C3$ hold. Let $(x^{(k)})_{k\in\mathbb{N}} \subset \ell^2$ converge weakly to $x \in \ell^2$ such that $\mathcal{R}(x^{(k)}) \to \mathcal{R}(x) < \infty$. Then $\|x^{(k)} - x\|_{\ell^2} \to 0$.*

Proof. We only consider the case $p > 1$ and $q < +\infty$. The proof for $p = 1$ and $q = +\infty$ is similar.

Let $\varepsilon > 0$. There exists a finite set $\Gamma \subset \Lambda$ such that

$$\sum_{\lambda\notin\Gamma} w_\lambda \phi(x_\lambda) \leq \varepsilon\,, \qquad \text{and} \qquad \sum_{\lambda\notin\Gamma} |x_\lambda|^2 \leq \varepsilon\ .$$

Since $x_\lambda^{(k)} \to x_\lambda$ for every $\lambda \in \Lambda$ and Γ is finite, there exists some $k_0 \in \mathbb{N}$ such that

$$\sum_{\lambda\in\Gamma} |x_\lambda^{(k)} - x_\lambda|^2 \le \varepsilon$$

for every $k \ge k_0$. Since ϕ is lower semi-continuous, there exists $k_1 \ge k_0$ such that

$$\sum_{\lambda\in\Gamma} w_\lambda \phi(x_\lambda^{(k)}) \ge \sum_{\lambda\in\Gamma} w_\lambda \phi(x_\lambda) - \varepsilon \ge \mathcal{R}(x) - 2\varepsilon$$

for every $k \ge k_1$. Conversely, the assumption that $\mathcal{R}(x^{(k)}) \to \mathcal{R}(x)$ implies the existence of $k_2 \ge k_1$ such that

$$\mathcal{R}(x^{(k)}) \le \mathcal{R}(x) + \varepsilon$$

for every $k \ge k_2$. Thus

$$\sum_{\lambda\notin\Gamma} w_\lambda \phi(x_\lambda^{(k)}) = \mathcal{R}(x_\lambda^{(k)}) - \sum_{\lambda\in\Gamma} w_\lambda \phi(x_\lambda^{(k)}) \le \mathcal{R}(x) + \varepsilon - (\mathcal{R}(x) - 2\varepsilon) = 3\varepsilon$$

for every $k \ge k_2$. In particular, we have for every $k \ge k_2$ and $\lambda \notin \Gamma$ that

$$3\varepsilon \ge w_\lambda \phi(x_\lambda^{(k)}) \ge C w_\lambda \frac{(x_\lambda^{(k)})^{2p}}{1 + (x_\lambda^{(k)})^{2p}}\,,$$

and therefore

$$(x_\lambda^{(k)})^{2p} \le \frac{3\varepsilon}{C \inf_\lambda w_\lambda - 3\varepsilon} =: K_\varepsilon\,.$$

Consequently, the reverse Hölder inequality implies that

$$\begin{aligned}
3\varepsilon \ge \sum_{\lambda\notin\Gamma} w_\lambda \phi(x_\lambda^{(k)}) \ge C \sum_{\lambda\notin\Gamma} w_\lambda \frac{(x_\lambda^{(k)})^{2p}}{1 + (x_\lambda^{(k)})^{2p}} &\ge \frac{C}{1+K_\varepsilon} \sum_{\lambda\notin\Gamma} w_\lambda |x_\lambda^{(k)}|^{2p} \\
&\ge \frac{C}{1+K_\varepsilon} \Big(\sum_{\lambda\in\Lambda} w_\lambda^{-q}\Big)^{-1/q} \Big(\sum_{\lambda\notin\Gamma} (x_\lambda^{(k)})^2\Big)^p
\end{aligned}$$

for every $k \ge k_2$, and thus

$$\sum_{\lambda\notin\Gamma} (x_\lambda^{(k)})^2 \le \Big(\frac{3(1+K_\varepsilon)}{C}\Big)^{1/p} \Big(\sum_{\lambda\in\Lambda} w_\lambda^{-q}\Big)^{1/pq} \varepsilon^{1/p} =: K'_\varepsilon \varepsilon^{1/p}\,.$$

Summarising the above estimates, we obtain that

$$\|x^{(k)} - x\|_{\ell^2}^2 \le \sum_{\lambda\in\Gamma} |x_\lambda^{(k)} - x_\lambda|^2 + 2\sum_{\lambda\notin\Gamma} |x_\lambda|^2 + 2\sum_{\lambda\notin\Gamma} |x_\lambda^{(k)}|^2 \le 3\varepsilon + 2K'_\varepsilon \varepsilon^{1/p}$$

for every $k \ge k_2$. Since K'_ε tends to zero as $\varepsilon \to 0$, the assertion follows. $\square$

Remark 3.8. The proofs in this section have made no use of the Hilbert space structure of ℓ^2. Indeed, each result can be formulated analogously for functionals on ℓ^r with $1 \le r < +\infty$ by simply replacing every occurence of the exponent 2 by r. In particular, the inequality (2) would read as

$$\phi(t) \ge \frac{C|t|^{rp}}{1+|t|^{rp}} \qquad \text{for every } t \in \mathbb{R} .$$

The same holds true for the results in Sections 4 and 5. ■

4 Well-posedness

Now we consider the regularising properties of the functional $\mathcal{T}_\alpha$ with $\mathcal{R}$ satisfying the conditions $C1$–$C3$. These results are a consequence of the Radon–Riesz property and the weak lower semi-continuity and weak coercivity of $\mathcal{R}$. Instead of providing complete proofs, only references to [17] are given. In addition, we show that the stronger growth condition $C3'$ implies the sparsity of every minimiser of $\mathcal{T}_\alpha(\cdot, y)$.

Strictly speaking, the results in [17] do not apply, as there the convexity of the regularisation term $\mathcal{R}$ is assumed. Also, the stability theorem in [17] does not consider varying regularisation parameters. An inspection of the proofs, however, shows that the assumption of convexity is only needed for the derivation of convergence rates and that the stability proof still holds if also the regularisation parameter is perturbed.

Proposition 4.1 (Existence). *Assume that the conditions $C1$–$C3$ hold. For every $\alpha > 0$ and $y \in Y$ the functional $\mathcal{T}_\alpha(\cdot, y)$ has a minimiser.*

Proof. See [17, Thm. 3.22]. □

Proposition 4.2 (Stability). *Assume that the conditions $C1$–$C3$ hold. Let $\alpha^{(k)} \to \alpha > 0$ and $y^{(k)} \to y \in Y$. Then every sequence*

$$x^{(k)} \in \arg\min\{\mathcal{T}_{\alpha^{(k)}}(x, y^{(k)}) : x \in \ell^2\}$$

has a subsequence $(x^{(k_l)})_{l\in\mathbb{N}}$ converging to a minimiser x_α of $\mathcal{T}_\alpha(\cdot, y)$ such that $\mathcal{R}(x^{(k_l)}) \to \mathcal{R}(x_\alpha)$. If the minimiser x_α is unique, then $x^{(k)} \to x_\alpha$.

Proof. Following the proof of [17, Thm. 3.23], we obtain a subsequence $(x^{(k_l)})_{l\in\mathbb{N}}$ converging to x_α such that $\mathcal{R}(x^{(k_l)}) \to \mathcal{R}(x)$. The norm convergence of the sequence then follows from Proposition 3.7. □

Proposition 4.3 (Convergence). *Assume that the conditions $C1$–$C3$ hold. Let $\alpha^{(k)} \to 0$ and $y^{(k)} \to y \in Y$ such that*

$$\frac{\|y^{(k)} - y\|^2}{\alpha^{(k)}} \to 0 .$$

Assume that there exists $x \in \operatorname{Dom} \mathcal{R}$ with $Ax = y$. Then every sequence

$$x^{(k)} \in \arg\min\bigl\{\mathcal{T}_{\alpha^{(k)}}(x, y^{(k)}) : x \in \ell^2\bigr\}$$

has a subsequence $(x^{(k_l)})_{l\in\mathbb{N}}$ converging to an $\mathcal{R}$-minimising solution $x^\dagger$ of the equation $Ax^\dagger = y$ such that $\mathcal{R}(x^{(k_l)}) \to \mathcal{R}(x^\dagger)$. If the $\mathcal{R}$-minimising solution $x^\dagger$ is unique, then $x^{(k)} \to x^\dagger$.

Proof. The weak convergence of a subsequence $(x^{(k_l)})_{l\in\mathbb{N}}$ to $x^\dagger$ together with the convergence of $\bigl(\mathcal{R}(x^{(k_l)})\bigr)_{l\in\mathbb{N}}$ to $\mathcal{R}(x^\dagger)$ follows from [17, Thm. 3.26]. The strong convergence of this sequence then follows from Proposition 3.7. □

Corollary 4.4. *Assume that the conditions C1–C3 hold. Let $y \in Y$ be such that the equation $Ax = y$ admits a unique $\mathcal{R}$-minimising solution $x^\dagger$. Define the function $H\colon \mathbb{R}_{>0} \times \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$,*

$$H(\alpha, \delta) := \sup\Bigl\{\|x_\alpha^\delta - x^\dagger\|_{\ell^2} : x_\alpha^\delta \in \arg\min_x \mathcal{T}_\alpha(x, y^\delta),\ \|y^\delta - y\| \leq \delta\Bigr\}\ .$$

Then for every $\varepsilon > 0$ there exists $\gamma > 0$ such that $H(\alpha, \delta) < \varepsilon$ whenever $0 < \alpha < \gamma$ and $0 < \delta^2 < \alpha\gamma$.

Proof. Assume to the contrary that there exists $\varepsilon > 0$ such that for every $k \in \mathbb{N}$ there exist $0 < \alpha^{(k)} < 1/k$ and $0 < (\delta^{(k)})^2 < \alpha^{(k)}/k$ such that $H(\alpha^{(k)}, \delta^{(k)}) \geq \varepsilon$. Then the definition of H implies that there exist sequences $(y^{(k)})_{k\in\mathbb{N}}$ with $\|y^{(k)} - y\| \leq \delta^{(k)}$, and $x^{(k)} \in \arg\min_x \mathcal{T}_{\alpha^{(k)}}(x, y^{(k)})$ such that $\|x^{(k)} - x^\dagger\|_{\ell^2} \geq \varepsilon/2$ for all $k \in \mathbb{N}$. In particular, the sequence $(x^{(k)})_{k\in\mathbb{N}}$ has no subsequence converging to $x^\dagger$, which contradicts Proposition 4.3. □

Proposition 4.5 (Sparsity). *Assume that the conditions C1, C2, and C3′ hold. Let $\alpha > 0$, $y \in Y$, and $x \in \arg\min\bigl\{\mathcal{T}_\alpha(x, y) : x \in \ell^2\bigr\}$. Then x is sparse.*

Proof. Define for $\mu \in \Lambda$ the sequence $\hat{x}^{(\mu)} := x - x_\mu e_\mu$. Since x minimises $\mathcal{T}_\alpha(\cdot, y)$, it follows that

$$\begin{aligned}\|Ax - y\|^2 + \alpha \sum_\lambda w_\lambda \phi(x_\lambda) &= \mathcal{T}_\alpha(x, y)\\ &\leq \mathcal{T}_\alpha(\hat{x}^{(\mu)}, y) = \|Ax - y - x_\mu A e_\mu\|^2 + \alpha \sum_{\lambda\neq\mu} w_\lambda \phi(x_\lambda)\ .\end{aligned}$$

Consequently,

$$\alpha w_\mu \phi(x_\mu) \leq x_\mu^2 \|Ae_\mu\|^2 + 2x_\mu \langle Ae_\mu, Ax - y\rangle \leq x_\mu^2 \|A\|^2 + 2x_\mu \langle e_\mu, A^*(Ax - y)\rangle$$

for every $\mu \in \Lambda$. With the estimate

$$C\alpha w_\mu \frac{|x_\mu|}{1 + |x_\mu|} \leq \alpha w_\mu \phi(x_\mu)$$

we obtain therefore that

$$|x_\mu| \le \frac{(1+\|x\|_{\ell^2})\big(x_\mu\|A\|^2 + 2\langle e_\mu, A^*(Ax-y)\rangle\big)}{C\alpha \inf_\lambda w_\lambda} x_\mu =: K_\mu x_\mu \,.$$

for every $\mu \in \Lambda$. Since $x \in \ell^2$ and $A^*(Ax-y) \in \ell^2$, it follows that the set $\Lambda' := \big\{\mu \in \Lambda : |K_\mu| \ge 1\big\}$ is finite. Since $x_\mu = 0$ whenever $\mu \notin \Lambda'$, this proves that x is sparse. □

5 Linear Convergence

Finally, we show that, under certain additional assumptions, the strongest growth condition at zero, $C4$, implies the linear convergence of minimisers x_α^δ to $x^\dagger$. The proof of this result closely resembles the proof of [12, Prop. 6.11], where the same convergence rate has been derived for constrained ℓ^p regularisation with $0 < p < 1$.

Theorem 5.1 (Linear Convergence). *Assume that conditions $C1$–$C4$ hold. Let $y \in Y$ be such that the equation $Ax = y$ admits a unique $\mathcal{R}$-minimising solution $x^\dagger$. Assume that $\Omega := \operatorname{supp}(x^\dagger)$ is finite and that the restriction of A to $\ell^2(\Omega)$ is injective. Define the function $H\colon \mathbb{R}_{>0} \times \mathbb{R}_{>0} \to \mathbb{R}_{\ge 0}$,*

$$H(\alpha,\delta) := \sup\Big\{\|x_\alpha^\delta - x^\dagger\|_{\ell^2} : x_\alpha^\delta \in \arg\min_x \mathcal{T}_\alpha(x,y^\delta),\ \|y^\delta - y\| \le \delta\Big\}\,.$$

Then there exist constants β_1, $\beta_2 > 0$ such that

$$H(\alpha,\delta) \le \frac{\beta_1\delta^2}{\alpha} + \beta_2\delta + \frac{\beta_2^2\alpha}{4\beta_1}$$

whenever $\alpha > 0$ and $\delta^2/\alpha > 0$ are small enough.

Proof. Denote by π_Ω, $\pi_\Omega^\perp$ the projections

$$\pi_\Omega x = \sum_{\lambda\in\Omega} x_\lambda e_\lambda\,, \qquad \pi_\Omega^\perp x = \sum_{\lambda\notin\Omega} x_\lambda e_\lambda\,.$$

As in the proofs of [11, Thm. 14, Thm. 15] one can prove the existence of $C_1 > 0$ such that

$$\|x - x^\dagger\|_{\ell^2} \le C_1\|A(x-x^\dagger)\| + \big(1 + C\|A\|\big)\|\pi_\Omega^\perp x\|_{\ell^2} \tag{4}$$

for every $x \in \ell^2$.

Since by assumption $D_+\phi(t) > -\infty$ and $D_-\phi(t) < +\infty$ for every $t \in \mathbb{R}$ and ϕ is bounded below by zero, there exists for every $t \in \mathbb{R}$ some $C(t) > 0$ such that

$$\phi(s) - \phi(t) \ge -C(t)|t-s|$$

for every $s \in \mathbb{R}$. Now define for $\sigma \in \{\pm 1\}^\Omega$ the sequence $\zeta(\sigma) \in \ell^2(\Omega)$ by $\zeta(\sigma)_\lambda = \operatorname{sgn}(\sigma_\lambda) w_\lambda C(x_\lambda^\dagger)$. Then

$$w_\lambda \phi(x_\lambda^\dagger) - w_\lambda \phi(t) \leq \zeta(\sigma)_\lambda (t - x_\lambda^\dagger)$$

for every $\lambda \in \Omega$ and $t \in \mathbb{R}$ with $\sigma_\lambda = \operatorname{sgn}(t - x_\lambda^\dagger)$. In particular,

$$\max_{\sigma \in \{\pm 1\}^\Omega} \left| \langle \zeta(\sigma), \pi_\Omega x - \pi_\Omega x^\dagger \rangle \right| \geq \mathcal{R}(x^\dagger) - \mathcal{R}(\pi x) \tag{5}$$

for every $x \in \ell^2$.

Denote now by $i_\Omega \colon \ell^2(\Omega) \to \ell^2(\Lambda)$ the embedding $i_\Omega x = x$. Then by assumption $A \circ i_\Omega \colon \ell^2(\Omega) \to \ell^2(\Lambda)$ is injective. Thus $(A \circ i_\Omega)^* = \pi_\Omega \circ A^* \colon \ell^2(\Lambda) \to \ell^2(\Omega)$ is surjective (see [19, Cor. VII.5.2]). In particular, $\zeta(\sigma) \in \operatorname{Range}(\pi_\Omega \circ A^*)$ for every $\sigma \in \{\pm 1\}^\Omega$. Hence there exists for every $\sigma \in \{\pm 1\}^\Omega$ some $\omega(\sigma) \in Y$ such that $\pi_\Omega \circ A^* \omega(\sigma) = \zeta(\sigma)$. Denote now

$$C_2 := \max_{\sigma \in \{\pm 1\}^\Omega} \|\omega(\sigma)\|_Y \ .$$

Then

$$\begin{aligned} \left| \langle \zeta(\sigma), \pi_\Omega x - \pi_\Omega x^\dagger \rangle \right| &\leq \left| \langle A^* \omega(\sigma), x - x^\dagger \rangle \right| + \left| \langle A^* \omega(\sigma), \pi_\Omega^\perp x \rangle \right| \\ &\leq C_2 \|A(x - x^\dagger)\| + C_2 \|A\| \, \|\pi_\Omega^\perp x\|_{\ell^2} \ . \end{aligned}$$

Consequently, using (5),

$$C_2 \|A(x - x^\dagger)\| \geq \mathcal{R}(x^\dagger) - \mathcal{R}(\pi x) - C_2 \|A\| \, \|\pi_\Omega^\perp x\|_{\ell^2} \ . \tag{6}$$

Since by assumption $D_+\phi(0) = +\infty$ and $D_-\phi(0) = -\infty$, there exists $\varepsilon > 0$ such that

$$\big(C_2 \|A\| + 1\big) |t| \leq w_\lambda \phi(t) \qquad \text{whenever } |t| \leq \varepsilon \ .$$

Thus we have for every $x \in \ell^2$ with $\|\pi_\Omega^\perp x\|_{\ell^\infty} \leq \varepsilon$ that

$$\begin{aligned} \big(C_2 \|A\| + 1\big) \|\pi_\Omega^\perp\|_{\ell^2} &\leq \big(C_2 \|A\| + 1\big) \|\pi_\Omega^\perp\|_{\ell^1} \\ &= \sum_{\lambda \notin \Omega} \big(C_2 \|A\| + 1\big) |x_\lambda| \leq \sum_{\lambda \notin \Omega} w_\lambda \phi(x_\lambda) = \mathcal{R}(\pi_\Omega^\perp x) \ . \end{aligned}$$

With (6) we therefore we obtain for every $x \in \ell^2$ with $\|\pi_\Omega^\perp x\|_{\ell^\infty} \leq \varepsilon$ the estimate

$$\|\pi_\Omega^\perp x\|_{\ell^2} \leq \mathcal{R}(\pi_\Omega^\perp x) - C_2 \|A\| \, \|\pi_\Omega^\perp x\|_{\ell^2} \leq \mathcal{R}(x) - \mathcal{R}(x^\dagger) + C_2 \|A(x - x^\dagger)\| \,,$$

and thus, using (4),

$$\|x - x^\dagger\|_{\ell^2} \leq (1 + C_1 \|A\|) \big(\mathcal{R}(x) - \mathcal{R}(x^\dagger)\big) + \big(C_1 + (1 + C_1 \|A\|) C_2\big) \|A(x - x^\dagger)\| \ . \tag{7}$$

Define now

$$\beta_1 := 1 + C_1 \|A\| \,, \qquad \beta_2 := C_1 + (1 + C_1 \|A\|) C_2 \ . \tag{8}$$

From Corollary 4.4 it follows that there exists $\gamma > 0$ such that $H(\alpha, \delta) < \varepsilon$ whenever $0 < \alpha < \gamma$ and $0 < \delta^2 < \alpha\gamma$. Let these constraints hold, let $y^\delta \in Y$ satisfy $\|y - y^\delta\| \leq \delta$, and choose some $x_\alpha^\delta \in \arg\min_x \mathcal{T}_\alpha(x, y^\delta)$. Then

$$\|Ax_\alpha^\delta - y^\delta\|^2 + \alpha\mathcal{R}(x_\alpha^\delta) \leq \|Ax^\dagger - y^\delta\|^2 + \alpha\mathcal{R}(x^\dagger) \leq \delta^2 + \alpha\mathcal{R}(x^\dagger)\,,$$

and thus

$$\mathcal{R}(x_\alpha^\delta) - \mathcal{R}(x^\dagger) \leq \frac{\delta^2 - \|Ax_\alpha^\delta - y^\delta\|^2}{\alpha}\,.$$

Since $\|x_\alpha^\delta - x^\dagger\|_{\ell^2} \leq H(\alpha, \delta) < \varepsilon$, we obtain using (7) and (8) that

$$\begin{aligned}
\|x_\alpha^\delta - x^\dagger\| &\leq \beta_1\bigl(\mathcal{R}(x_\alpha^\delta) - \mathcal{R}(x^\dagger)\bigr) + \beta_2\|Ax_\alpha^\delta - y\| \\
&\leq \frac{\beta_1\delta^2}{\alpha} - \frac{\beta_1\|Ax_\alpha^\delta - y^\delta\|^2}{\alpha} + \beta_2\|Ax_\alpha^\delta - y^\delta\| + \beta_2\delta \\
&\leq \frac{\beta_1\delta^2}{\alpha} + \beta_2\delta + \frac{\beta_2^2\alpha}{4\beta_1}\,,
\end{aligned}$$

which proves the assertion. □

6 Summary

In this paper, we have studied Tikhonov regularisation on ℓ^2 with general weighted penalty terms of the form $\mathcal{R}(x) = \sum_\lambda w_\lambda\phi(x_\lambda)$. Fairly general requirements have been given that guarantee the well-posedness of the regularisation method. Moreover, under an additional boundedness assumption for the chosen weights, these requirements have been shown to be necessary for the weak lower semi-continuity and weak coercivity of the regularisation term. In particular, these conditions encompass weighted ℓ^p regularisation with $0 < p \leq 2$, but also ℓ^0 regularisation with additional hard constraints.

A central focus of this paper lies on the possible application of the considered regularisation method to the recovery of sparse sequences. We have formulated a sufficient growth condition for ϕ at zero that enforces the minimisers of the Tikhonov functional $\mathcal{T}_\alpha$ to be sparse. In addition, we have treated the question of convergence rates. Here we have assumed that the unique $\mathcal{R}$-minimising solution $x^\dagger$ of $Ax = y$ is sparse and that A satisfies a kind of finite basis injectivity property. Requiring that ϕ has a superlinear growth at zero, we have shown that the minimisers of $\mathcal{T}_\alpha(\cdot, y^\delta)$ converge linearly to $x^\dagger$ as $\alpha \sim \|y^\delta - y\| \to 0$. At the moment, these are the weakest conditions on $x^\dagger$ and A, under which a linear convergence rate has been derived.

Acknowledgement

This work has been supported by the Austrian Science Fund (FWF) within the national research network *Industrial Geometry*, project 9203-N12.

References

[1] K. Bredies and D. Lorenz. Regularization with non-convex separable constraints. Technical Report 13, DFG-Schwerpunktprogramm 1324, 2009.

[2] A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev.*, 51(1):34–81, 2009.

[3] 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.

[4] A. Chambolle, R. A. DeVore, N. Lee, and B. J. Lucier. Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Trans. Image Process.*, 7(3):319–335, 1998.

[5] P. L. Combettes and J.-C. Pesquet. Proximal thresholding algorithm for minimization over orthonormal bases. *SIAM J. Optim.*, 18(4):1351–1376, 2007.

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

[7] 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.

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

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

[10] M. Grasmair. Well-posedness and convergence rates for sparse regularization with sublinear l^q penalty term. Reports of FSP S092 - "Industrial Geometry" 74, University of Innsbruck, Austria, 2008. Accepted for publication in Inverse Probl. Imaging.

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

[12] 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.

[13] E. Hewitt and K. Stromberg. *Real and Abstract Analysis.* Springer Verlag, New York, 1965.

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

[15] R. E. Megginson. *An Introduction to Banach Space Theory*, volume 183 of *Graduate Texts in Mathematics*. Springer Verlag, New York, 1998.

[16] R. Ramlau and C. Zarzer. On the optimization of a tikhonov functional with non-convex sparsity constraints. Technical report, Johann Radon Institute for Computational and Applied Mathematics (RICAM), 2009.

[17] 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.

[18] J. A. Tropp. Just relax: convex programming methods for identifying sparse signals in noise. *IEEE Trans. Inf. Theory*, 52(3):1030–1051, 2006.

[19] K. Yosida. *Functional Analysis*, volume 123 of *Die Grundlehren der Mathematischen Wissenschaften*. Academic Press Inc., New York, 1965.

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