COMMUN. MATH. SCI.
Vol. 9, No. 4, pp. 1129–1141

WELL-POSEDNESS CLASSES FOR SPARSE REGULARIZATION*

MARKUS GRASMAIR†

Abstract. Because of their sparsity enhancing properties, ℓ^1 penalty terms have recently received much attention in the field of inverse problems. Also, it has been shown that certain properties of the linear operator A to be inverted imply that ℓ^1-regularization is equivalent to ℓ^0-regularization, which tries to minimise the number of non-zero coefficients. In the context of compressed sensing, one usually assumes a restricted isometry property, which requires that the operator A acts almost like an isometry on certain low dimensional sub-spaces. In this paper, we show that similar properties appear naturally when one studies the question of well-posedness of ℓ^0-regularization. Moreover, we derive a complete characterisation of those linear operators A for which ℓ^0-regularization is well-posed. It turns out that neither boundedness nor invertibility of A are necessary conditions; compact operators, however, are shown not to be suited for ℓ^0-regularization.

Key words. Sparsity, quasi-solutions, restricted isometry.

MSC. 65J20, 47A52.

1. Introduction

While the basic assumption of classical regularization methods for the stable inversion of ill-posed operator equations is either boundedness or smoothness of the solution, the theory of sparse regularization is based on the assumption that the true solution has a finite expansion with respect to some given basis of the space of definition. This sparsity of the solution can be enforced by employing the number of non-zero coefficients as a regularization term. There are two major problems with this approach: The first problem is the fact that, in general, this does not yield a well-posed regularization method, as has been observed in [15]; the number of non-zero coefficients is no coercive functional, and therefore the regularization functional need not attain its minimum. The second difficulty lies in the actual computation of the solution, provided it exists.

In order to obtain a problem that is computationally tractable, it has been suggested in [6] to use the ℓ^1-norm as a sparsity enforcing regularization term. Then one obtains a convex minimisation problem that can be solved by standard methods. This approach was rigorously justified later, when it was shown that, under certain assumptions, the ℓ^1-minimiser is also the sparsest solution (see, for instance, [4, 5, 8, 9]). In [5], the main assumption is that the operator A to be inverted satisfies a certain *restricted isometry property*, which requires that A acts almost like an orthogonal operator on the class of all sufficiently sparse vectors (see also [3]).

In [7, 12, 15, 17, 18], the method of ℓ^1-regularization has been considered from an inverse problems point of view. It has been shown there that ℓ^1-regularization provides a well-posed regularization method and, in addition, that it may have exceedingly good properties: If the true solution of the considered equation is sparse and satisfies a range condition, and the operator A satisfies a certain *restricted injectivity property*, then the regularized solution converges linearly to the true solution as the noise level decreases to zero. In [13], the injectivity condition has been replaced by the assumption of uniqueness of the ℓ^1-minimising solution. In addition, it has been

*Received: October 13, 2010; accepted (in revised version): May 5, 2011. Communicated by Martin Burger.
†Computational Science Center, University of Vienna, Nordbergstr. 15, A–1090 Vienna, Austria (markus.grasmair@univie.ac.at).

shown that the restricted isometry property implies a uniform range condition on the set of all sufficiently sparse elements.

There are, however, fundamental differences between the latter results and those on compressed sensing. First, the convergence rates derived from restricted isometry properties hold uniformly on the set of sufficiently sparse vectors, while those in [13] depend strongly on the solution. Second, the results on convergence rates for ℓ^1-regularization in [13] make no assertion concerning the question whether the ℓ^1-minimising solutions are actually the sparsest ones (note however [16], where sparsity results for ℓ^1-regularization were derived using methods from compressed sensing applied to the setting of inverse problems). Indeed, the range condition postulated in [13] seems not be strong enough to guarantee such a sparsity. On the other hand, it might be possible to substantially weaken the assumptions in [5] and still be able to obtain this equivalence of ℓ^1- and ℓ^0-regularization. Also, it might be possible that weaker assumptions would at least imply the equivalence of ℓ^0-regularization and ℓ^p-regularization for some $0<p<1$. Such regularization methods with non-convex penalty term have for instance been studied in [1, 10, 11, 21] (see also [15], where some preliminary first results on well-posedness of the non-convex case were presented).

If one aims for the derivation of more general equivalence results, it makes sense to study first the properties of ℓ^0-regularization more closely. In particular, because regularization with ℓ^p-penalty terms is well-posed for every $p>0$ (see [10]), one should derive conditions for the operator A that guarantee this well-posedness also for the case of an ℓ^0-penalty term. In this paper, we will approach this task by studying the method of quasisolutions, where one assumes that a strict bound for the penalty term is known. In the setting of sparse regularization, this means that an upper bound s for the number of non-zero coefficients of the expected true solutions is known a priori. The same assumption is also present in the theory of compressed sensing.

The main result of the paper is Theorem 3.1, where a characterisation of those linear operators is given for which the assumption of s-sparsity leads to a well-posed problem. It turns out that the conditions one obtains are closely connected to the restricted isometry property, though far less restrictive. As a consequence of these conditions, we obtain that compact operators are only in the finite dimensional case susceptible to ℓ^0-regularization. Still, the class of operators for which s-sparsity is a meaningful regularization assumption contains more operators than only isomorphisms: We show by means of two explicit examples that neither the boundedness of the operator A nor the closedness of the range of A are necessary for ℓ^0-regularization to be well-posed.

2. Sparse regularization

Let Λ be some countable index set, Y some Hilbert space, and $A\colon \ell^2(\Lambda)\to Y$ a linear operator. The goal is to solve, for given data $y\in Y$, the operator equation

$$Ax=y. \tag{2.1}$$

If the operator A is ill-posed, then solving this equation, if possible, in general does not yield any meaningful results, as small errors in the data y can lead to arbitrarily large errors in the solution. In order to obtain useful results nevertheless, it is necessary to have some additional a priori knowledge about the solution of the equation and to use it in some approximate solution process. One possibility for such a priori knowledge is sparsity, where one assumes that the support of the solution $x^\dagger=(x^\dagger_\lambda)_{\lambda\in\Lambda}$ of (2.1), that is, the set

$$\operatorname{supp}(x^\dagger)=\{\lambda\in\Lambda : x^\dagger_\lambda\neq 0\},$$

is a finite set.

There are several methods for exploiting this knowledge. If one not only knowns that $\mathrm{supp}(x^\dagger)$ is finite, but an additional estimate of the data error in the form $\|y - y^\delta\| \le \delta$ is available, then one can solve the constrained minimisation problem

$$|\mathrm{supp}(x)| \to \min \qquad \text{subject to } \|Ax - y^\delta\|^2 \le \delta^2. \tag{2.2}$$

Also, it is possible to apply a Tikhonov type of regularization, and solve the unconstrained minimisation problem

$$\|Ax - y^\delta\|^2 + \alpha |\mathrm{supp}(x)| \to \min, \tag{2.3}$$

where the regularization parameter $\alpha > 0$ is chosen in some suitable manner. If, on the other hand, an explicit bound for $|\mathrm{supp}(x^\dagger)|$ is given, for instance the knowledge that $|\mathrm{supp}(x^\dagger)| \le s$, then it makes sense to compute the *quasi-solution* of (2.1), which is defined as

$$x^{(s)} := \arg\min\big\{ \|Ax - y^\delta\|^2 : x \in X_s \big\}, \tag{2.4}$$

where

$$X_s := \big\{ x \in \ell^2(\Lambda) : |\mathrm{supp}(x)| \le s \big\}.$$

The problem with all of the three models (2.2), (2.3), and (2.4) is that, in general, none of them is well-defined. The reason is that, while the mapping $x \mapsto |\mathrm{supp}(x)|$ is weakly lower semi-continuous, it is not coercive (see [11]). Thus, direct methods for proving the existence of solutions can only be applied, if the non-coercivity of $|\mathrm{supp}(x)|$ is compensated by the coercivity of the fidelity term $\|Ax - y^\delta\|^2$ on the set X_s. To see what can happen in the general case, we consider a simple example, which makes use of the same construction as an example in [15, Section 5.2], where it is shown that Tikhonov regularization with an ℓ^0-penalty term need not be well-posed.

Before that we clarify some notation that is used throughout the text. Whenever Λ is some (countable) index set and $x \in \ell^2(\Lambda)$, we denote by $x_\lambda \in \mathbb{R}$, $\lambda \in \Lambda$, the λ-th coefficient of x. Moreover, we denote by $e^\lambda \in \ell^2(\Lambda)$, $\lambda \in \Lambda$, the standard basis vector with coefficients $(e^\lambda)_\lambda = 1$ and $(e^\lambda)_\mu = 0$ for every $\mu \in \Lambda \setminus \{\lambda\}$. That is, for all $x \in \ell^2(\Lambda)$ we have the representation $x = \sum_{\lambda \in \Lambda} x_\lambda e^\lambda$.

EXAMPLE 2.1. *Define* $A\colon \ell^2(\mathbb{N}) \to \mathbb{R}^2$ *by*

$$Ae^k = \frac{\cos(k)e^1 + \sin(k)e^2}{k}.$$

Then

$$\begin{aligned}
\|Ax\|^2 &= \Big(\sum_k \frac{\cos(k)x_k}{k}\Big)^2 + \Big(\sum_k \frac{\sin(k)x_k}{k}\Big)^2 \\
&\le \Big(\sum_k \frac{\cos(k)^2}{k^2}\Big)\Big(\sum_k x_k^2\Big) + \Big(\sum_k \frac{\sin(k)^2}{k^2}\Big)\Big(\sum_k x_k^2\Big) \\
&\le \frac{\pi^2}{3}\|x\|^2,
\end{aligned}$$

showing that A *defines a bounded linear operator on* $\ell^2(\mathbb{N})$.

Now note that the set $A(X_1) = \{A(te^k) : t \in \mathbb{R},\ k \in \mathbb{N}\}$ *is dense in* $\mathbb{R}^2$. *Thus, whenever* $y^\delta \in \mathbb{R}^2 \setminus A(X_1)$ *is given, the problem of minimising* $\|Ax - y^\delta\|^2$ *over* X_1 *has no solution. Even more, the Tikhonov functional* (2.3) *attains no solution if* $\|y^\delta\|^2 > \alpha$. *Indeed, the density of* $A(X_1)$ *in* $\mathbb{R}^2$ *implies that* $\inf_{x \in X_1} \|Ax - y^\delta\|^2 = 0$ *and therefore*

$$\inf_{x \in \ell^2(\mathbb{N})} \bigl(\|Ax - y^\delta\|^2 + \alpha|\mathrm{supp}(x)|\bigr) \le \alpha.$$

On the other hand, with $x = 0$ *we have* $\|Ax - y^\delta\|^2 + \alpha|\mathrm{supp}(x)| = \|y^\delta\|^2 > \alpha$, *which implies that* $x = 0$ *is no minimiser. If, however,* $x \in X_1$ *were a minimiser, then* $\|Ax - y^\delta\|^2 = 0$, *contradicting the assumption that* $y^\delta \notin A(X_1)$.

In the following, we will concentrate on the concept of quasi-solutions, that is, the model (2.4). These types of models are classically treated within the concept of well-posedness classes [19]:

Definition 2.1. *Let* X *and* Y *be topological spaces and let* $A\colon X \to Y$. *The set* $\tilde X \subset X$ *is a* well-posedness class *for* A, *if the restriction of* A *to* $\tilde X$ *is well-posed in the sense of Hadamard. That is, the following conditions are satisfied:*

- *The restriction of* A *to* $\tilde X$ *is continuous.*
- *The restriction of* A *to* $\tilde X$ *is injective.*
- *The mapping* $A^{-1}\colon A(\tilde X) \to \tilde X$ *is continuous.*

In other words, the set $\tilde X$ is a well-posedness class for A if the operator A is a homeomorphism between $\tilde X$ and its image $A(\tilde X) \subset Y$.

In the following section, we will derive necessary and sufficient conditions for the linear operator A that guarantee that the sets X_s, for given $s \in \mathbb{N}$, are well-posedness classes for A.

3. Well-posedness classes

Let

$$\rho_s := \sup\left\{ \frac{\|Ax\|}{\|x\|} : x \in X_s \setminus \{0\} \right\},$$
$$\sigma_s := \inf\left\{ \frac{\|Ax\|}{\|x\|} : x \in X_s \setminus \{0\} \right\}.$$

Moreover, define for $x \in \ell^2(\Lambda)$ with $Ax \neq 0$

$$\tau_s(x) := \sup\left\{ \frac{\langle Ax, A\tilde x\rangle}{\|Ax\|\|A\tilde x\|} : \tilde x \in X_s,\ \mathrm{supp}(\tilde x) \cap \mathrm{supp}(x) = \emptyset,\ A\tilde x \neq 0 \right\},$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product on the Hilbert space Y, and define for $\Omega \subset \Lambda$

$$\tau_s(\Omega) := \sup\{\tau_s(x) : \mathrm{supp}(x) \subset \Omega,\ Ax \neq 0\}. \tag{3.1}$$

Here we set $\tau_s(\Omega) := 0$ if $Ax = 0$ for every x satisfying $\mathrm{supp}(x) \subset \Omega$. Note that the supremum in (3.1) is attained whenever $\Omega \subset \Lambda$ is a finite set and $Ax \neq 0$ for some x satisfying $\mathrm{supp}(x) \subset \Omega$. That is, whenever Ω is a finite set and $\tau_s(\Omega) > 0$, there exists $x \in \ell^2(\Lambda)$ with $\mathrm{supp}(x) \subset \Omega$ such that $\tau_s(\Omega) = \tau_s(x)$.

Theorem 3.1. *The set* X_s *is a well-posedness class for the operator* A *if and only if the following hold:*

$$\rho_s < \infty, \tag{3.2}$$
$$\sigma_s > 0, \tag{3.3}$$
$$\tau_s(\Omega) < 1 \quad \text{for every } \Omega \subset \Lambda \text{ with } |\Omega| \le s. \tag{3.4}$$

Proof. Assume first that X_s is a well-posedness class for A, that is, the restriction of A to X_s is continuous, injective, and has a continuous inverse. We show by contradiction that the conditions (3.2)–(3.4) are satisfied. Assume first that (3.2) does not hold. Then there exists a sequence $(x^{(k)})_{k\in\mathbb{N}}\subset X_s$ with $\|x^{(k)}\|\le 1$ for all k and $\|Ax^{(k)}\|\to\infty$. Setting $\tilde{x}^{(k)}:=x^{(k)}/\|Ax^{(k)}\|$, it follows that $\tilde{x}^{(k)}\to 0$ while $\|A\tilde{x}^{(k)}\|=1$ giving a contradiction to the continuity of A on X_s.

Now assume that (3.3) does not hold. Then there exists a sequence $(x^{(k)})_{k\in\mathbb{N}}\subset X_s$ such that $\|x^{(k)}\|=1$ for all k and $Ax^{(k)}\to 0$. Obviously, this contradicts the continuous invertibility of $A|_{X_s}$.

Finally assume that (3.2) holds, but (3.4) does not hold. Let $\Omega\subset\Lambda$ be such that $|\Omega|\le s$ and $\tau_s(\Omega)=1$. Then there exist $x\in\ell^2(\Lambda)$ with $\operatorname{supp}(x)\subset\Omega$ and $\|Ax\|=1$, and a sequence $(x^{(k)})_{k\in\mathbb{N}}\subset\ell^2(\Lambda)$ such that $x^{(k)}\in X_s$, $\operatorname{supp}(x^{(k)})\cap\operatorname{supp}(x)=\emptyset$, $\|Ax^{(k)}\|=1$ for all $k\in\mathbb{N}$, and

$$\frac{\langle Ax,Ax^{(k)}\rangle}{\|Ax\|\|Ax^{(k)}\|}=\langle Ax,Ax^{(k)}\rangle\to 1.$$

Then

$$\|A(x-x^{(k)})\|^2=\|Ax\|^2+\|Ax^{(k)}\|^2-2\langle Ax,Ax^{(k)}\rangle=2(1-\langle Ax,Ax^{(k)}\rangle)\to 0.$$

On the other hand (3.2) implies that $\|x\|\ge\|Ax\|/\rho_s=1/\rho_s$ and also $\|x^{(k)}\|\ge 1/\rho_s$ for all $k\in\mathbb{N}$. Because $\operatorname{supp}(x)\cap\operatorname{supp}(x^{(k)})=\emptyset$, this implies that $\|x-x^{(k)}\|\ge\sqrt{2}/\rho_s$, showing that $Ax^{(k)}$ converges to Ax while $x^{(k)}$ does not converge to x. This gives the necessary contradiction.

For the converse direction, assume that (3.2)–(3.4) hold. We first show that $A|_{X_s}$ is continuous. Let therefore $x\in X_s$ and assume that $(x^{(k)})_{k\in\mathbb{N}}\subset X_s$ converges to x. Denote $\pi_x\colon\ell^2(\Lambda)\to\ell^2(\Lambda)$ the projection on the subspace spanned by the basis elements in the support of x, that is,

$$\pi_x(\tilde{x})=\sum_{\lambda\in\operatorname{supp}(x)}\tilde{x}_\lambda e^\lambda,$$

and define $\pi_x^\perp:=\mathrm{Id}-\pi_x$. Then $x-\pi_x(x^{(k)})\in X_s$ for all k and, similarly, $\pi_x^\perp(x^{(k)})\in X_s$ for all k. Therefore,

$$\begin{aligned}\|A(x^{(k)}-x)\|&\le\|A(\pi_x(x^{(k)})-x)\|+\|A\pi_x^\perp(x^{(k)})\|\\&\le\rho_s\big(\|\pi_x(x^{(k)})-x\|+\|\pi_x^\perp(x^{(k)})\|\big)\le\sqrt{2}\rho_s\|x^{(k)}-x\|,\end{aligned}$$

proving the continuity of $A|_{X_s}$.

Now assume that $x\in X_s$ and $(x^{(k)})_{k\in\mathbb{N}}\subset X_s$ are such that $Ax^{(k)}\to Ax$. We have to show that also $x^{(k)}\to x$. Let $\Omega:=\operatorname{supp}(x)$. Then by assumption $\tau_s(\Omega)<1$. Moreover, we have $\operatorname{supp}(\pi_x(x^{(k)})-x)\subset\operatorname{supp}(x)\subset\Omega$, and therefore, as $|\operatorname{supp}(\pi_x^\perp(x^{(k)}))|\le s$ and $\operatorname{supp}(\pi_x^\perp(x^{(k)}))\cap\Omega=\emptyset$,

$$\langle A(\pi_x(x^{(k)})-x),A\pi_x^\perp(x^{(k)})\rangle\le\tau_s(\Omega)\|A(\pi_x(x^{(k)})-x)\|\|A\pi_x^\perp(x^{(k)})\|$$

for all k. Thus,

$$\begin{aligned}
&\|A(x^{(k)}-x)\|^2 \\
&\qquad = \|A(\pi_x(x^{(k)})-x)\|^2 + \|A\pi_x^{\perp}(x^{(k)})\|^2 - 2\langle A(\pi_x(x^{(k)})-x), A\pi_x^{\perp}(x^{(k)})\rangle \\
&\qquad \geq \|A(\pi_x(x^{(k)})-x)\|^2 + \|A\pi_x^{\perp}(x^{(k)})\|^2 \\
&\qquad\qquad - 2\tau_s(\Omega)\|A(\pi_x(x^{(k)})-x)\|\|A\pi_x^{\perp}(x^{(k)})\| \\
&\qquad = \big(\|A(\pi_x(x^{(k)})-x)\| - \|A\pi_x^{\perp}(x^{(k)})\|\big)^2 \\
&\qquad\qquad + 2(1-\tau_s(\Omega))\|A(\pi_x(x^{(k)})-x)\|\,\|A\pi_x^{\perp}(x^{(k)})\| \\
&\qquad \geq 0.
\end{aligned}$$

Since by assumption $\|A(x^{(k)}-x)\|$ converges to zero and $1-\tau_s(\Omega)>0$, it follows that so do the sequences $\|A(\pi_x(x^{(k)})-x)\|-\|A\pi_x^{\perp}(x^{(k)})\|$ and $\|A(\pi_x(x^{(k)})-x)\|\,\|A\pi_x^{\perp}(x^{(k)})\|$. Consequently, we obtain that

$$\|A(\pi_x(x^{(k)})-x)\| \to 0 \qquad \text{and} \qquad \|A\pi_x^{\perp}(x^{(k)})\| \to 0. \tag{3.5}$$

Because $\pi_x(x^{(k)})-x \in X_s$ and $\pi_x^{\perp}(x^{(k)}) \in X_s$ for all $k\in\mathbb{N}$, it follows from (3.3) that

$$\begin{aligned}
\|x^{(k)}-x\|^2 &= \|\pi_x(x^{(k)})-x\|^2 + \|\pi_x^{\perp}(x^{(k)})\|^2 \\
&\leq \sigma_s^2\big(\|A(\pi_x(x^{(k)})-x)\|^2 + \|A\pi_x^{\perp}(x^{(k)})\|^2\big).
\end{aligned}$$

Now the convergence of $(x^{(k)})$ to x follows from (3.5). ∎

COROLLARY 3.2. *Assume that X_1 is a set of well-posedness for the linear operator $A\colon \ell^2(\Lambda)\to Y$ and that Λ is an infinite set. Then A is non-compact.*

Proof. For ease of notation we assume without loss of generality that $\Lambda=\mathbb{N}$.

Assume to the contrary that A is compact and consider the sequence of basis vectors $(e^k)_{k\in\mathbb{N}}$. This sequence converges weakly to zero in $\ell^2(\mathbb{N})$. Moreover, the compactness of A implies in particular that A is bounded, and therefore also the sequence $(Ae^k)_{k\in\mathbb{N}}$ converges weakly to zero. Now the compactness of A implies that $(Ae^k)_{k\in\mathbb{N}}$ converges to zero with respect to the norm, and thus $\|Ae^k\|\to 0$. This, however, is a contradiction to the assumption that X_1 is a set of well-posedness for A, as Theorem 3.1 in particular implies that $\|Ae^k\|\geq\sigma_1>0$ for all $k\in\mathbb{N}$. ∎

REMARK 3.3. Most of the technicalities in the proof of Theorem 3.1 stem from the fact that the sets X_s are not linear sub-spaces of $\ell^2(\Lambda)$. If the sets X_s were linear spaces, then ρ_s would simply equal the norm of the restriction of A to X_s, and σ_s the reciprocal of the norm of its inverse; thus, the conditions $\rho_s<\infty$ and $\sigma_s>0$ alone would be equivalent to X_s being a well-posedness class for A. Here, because of the non-linearity, the additional condition (3.4) is needed. Note, however, that the proof of Theorem 3.1 shows that this condition is only required for the continuity of the inverse: the restriction of A to X_s is continuous, if and only if $\rho_s<\infty$.

REMARK 3.4. The (s,s')-*orthogonality constant* of A(see [5]) is defined as

$$\tau_{s,s'} := \sup\big\{\tau_s(\Omega) : \Omega\subset\Lambda,\ |\Omega|\leq s'\big\}.$$

It can be used in conditions that guarantee that the solution of constrained ℓ^1-regularisation is at the same time the sparsest solution of the equation $Ax=y$ (see [2] for a collection of results of that type). Obviously, the condition $\tau_{s,s}<1$ implies condition (3.4) of Theorem 3.1. Conversely, however, it is easily possible in the infinite dimensional case that (3.4) holds, although $\tau_{s,s}=1$.

REMARK 3.5. There is also a connection between τ_s and the notion of *coherence* of a linear mapping, which can be defined as

$$\mu=\sup_{i\neq j}\frac{|\langle Ae^i,Ae^j\rangle|}{\|Ae^i\|\|Ae^j\|}.$$

Coherence is one of the main tools in compressed sensing for deriving not only the equivalence of ℓ^1 and ℓ^0-regularization, but also the correctness of several greedy algorithms that can be used for solving the minimisation problem numerically (see [9] for a large collection of results). It is easy to see that $\mu(A)=\tau_{1,1}$. In [20], the *babel function* $\mu_1\colon\mathbb{N}\to\mathbb{R}_{\geq 0}$, defined by

$$\mu_1(m):=\sup_{\substack{\Omega\subset\Lambda\\|\Omega|=m}}\sup_{i\notin\Omega}\sum_{j\in\Omega}\frac{|\langle Ae^i,Ae^j\rangle|}{\|Ae^i\|\|Ae^j\|},$$

has been introduced in order to obtain sharper results on the validity of greedy algorithms. This original definition has been dissected in [14] in order to analyse different a priori assumptions on the support of $x^\dagger$ than that of s-sparsity. In the course of this analysis, the *setwise babel function* has been defined as the mapping that assigns to every index set $\Omega\subset\Lambda$ the number

$$\mu_1(\Omega):=\sup_{i\notin\Omega}\sum_{j\in\Omega}\frac{|\langle Ae^i,Ae^j\rangle|}{\|Ae^i\|\|Ae^j\|}.$$

This definition of $\mu_1(\Omega)$ is very close to that of $\tau_1(\Omega)$ in the present paper. The main difference to τ_1 is that μ_1 uses the sum of the absolute values of the inner products of Ae^i and Ae^j, which can be interpreted as an ℓ^1-norm on a sub-space of Y. Indeed, the arguments in [14] are based on an estimation of the norm of the operator A with respect to suitable ℓ^1-norms (see also [16], where similar arguments are used in the infinite dimensional case).

4. Examples

In this section we show by means of two concrete examples that the conditions in Theorem 3.1 imply neither boundedness nor bounded invertibility of the operator A, even if they are satisfied for every $s\in\mathbb{N}$. In the first example, we construct an unbounded operator, for which X_s is a set of well-posedness for each $s\in\mathbb{N}$.

EXAMPLE 4.1. *Consider the sets*

$$\Lambda:=\{(k,l):k,l\in\mathbb{N},\ 1\leq l\leq k\},$$
$$\Lambda':=\{(k,l):k,l\in\mathbb{N},\ 0\leq l\leq k\}.$$

Let moreover $A\colon\ell^2(\Lambda)\to\ell^2(\Lambda')$ be any linear operator satisfying

$$Ae^{k,l}=e^{k,0}+e^{k,l}\qquad\text{for } (k,l)\in\Lambda.$$

In the following, we show that the operator A is unbounded, but that every set X_s is a set of well-posedness for A.

In order to see that A is unbounded, consider $x^{(k)} \subset \ell^2(\Lambda)$ defined as $x^{(k)} := \sum_{1\le l\le k} e^{k,l}$. Then $Ax^{(k)} = ke^{k,0} + \sum_{1\le l\le k} e^{k,l}$, and therefore

$$\|x^{(k)}\|^2 = k, \qquad \textit{while} \qquad \|Ax^{(k)}\|^2 = k^2 + k. \tag{4.1}$$

Now let $s\in\mathbb{N}$ be fixed. In order to show that the set X_s is a set of well-posedness for A, we have to verify that $\rho_s<\infty$, $\sigma_s>0$, and $\tau_s(\Omega)<1$ for every $\Omega\subset\Lambda$ with $|\Omega|\le s$. Even more, we will show that the (s,s)-orthogonality constant $\tau_{s,s}$ is strictly smaller than 1.

Assume first that $x=\sum_k\sum_{1\le l\le k} x_{k,l}e^{k,l}\in X_s$. Then

$$\|Ax\|^2 = \sum_k \left(\sum_{1\le l\le k} x_{k,l}^2 + \left[\sum_{1\le l\le k} x_{k,l} \right]^2 \right). \tag{4.2}$$

This immediately shows that $\|Ax\|^2 \ge \|x\|^2$, implying that $\sigma_s \ge 1$. Moreover, because $x\in X_s$, it follows that there exist at most s pairs (k,l) such that $x_{k,l}\neq 0$. Consequently, making use of the estimate $(y_1+\ldots+y_s)^2 \le s(y_1^2+\ldots+y_s^2)$, we obtain that

$$\|Ax\|^2 \le \sum_k \sum_{1\le l\le k} (1+s)x_{k,l}^2 = (1+s)\|x\|^2,$$

showing that $\rho_s\le\sqrt{1+s}$. Together with (4.1) *we obtain that, in fact, we have equality; that is, $\rho_s=\sqrt{1+s}$.*

Now let $\tilde{x}=\sum_k\sum_{1\le l\le k}\tilde{x}_{k,l}e^{k,l}\in X_s$ be such that $\operatorname{supp}(x)\cap\operatorname{supp}(\tilde{x})=\emptyset$. Then $x_{k,l}\tilde{x}_{k,l}=0$ for every $(k,l)\in\Lambda$, showing that

$$\begin{aligned}
\langle Ax, A\tilde{x}\rangle &= \sum_k \left(\sum_{1\le l\le k} x_{k,l}\tilde{x}_{k,l} + \left(\sum_{1\le l\le k} x_{k,l} \right) \left(\sum_{1\le l\le k} \tilde{x}_{k,l} \right) \right) \\
&= \sum_k \left(\left(\sum_{1\le l\le k} x_{k,l} \right) \left(\sum_{1\le l\le k} \tilde{x}_{k,l} \right) \right) \\
&\le \sum_k \left(\left| \sum_{1\le l\le k} x_{k,l} \right| \left| \sum_{1\le l\le k} \tilde{x}_{k,l} \right| \right).
\end{aligned}$$

Now note that the fact that $|\operatorname{supp}(x)|\le s$ implies the inequality

$$\left(\sum_{1\le l\le k} x_{k,l} \right)^2 \le s \sum_{1\le l\le k} x_{k,l}^2,$$

which in turn shows that

$$\left| \sum_{1\le l\le k} x_{k,l} \right| \le \sqrt{\frac{s}{s+1}} \sqrt{ \sum_{1\le l\le k} x_{k,l}^2 + \left(\sum_{1\le l\le k} x_{k,l} \right)^2 },$$

and, similarly,

$$\left|\sum_{1\le l\le k}\tilde{x}_{k,l}\right| \le \sqrt{\frac{s}{s+1}}\sqrt{\sum_{1\le l\le k}\tilde{x}_{k,l}^2+\left(\sum_{1\le l\le k}\tilde{x}_{k,l}\right)^2}.$$

Therefore,

$$\begin{aligned}\langle Ax,A\tilde{x}\rangle &\le \frac{s}{s+1}\sum_k\sqrt{\sum_{1\le l\le k}x_{k,l}^2+\left(\sum_{1\le l\le k}x_{k,l}\right)^2}\sqrt{\sum_{1\le l\le k}\tilde{x}_{k,l}^2+\left(\sum_{1\le l\le k}\tilde{x}_{k,l}\right)^2}\\ &\le \frac{s}{s+1}\sqrt{\sum_k\sum_{1\le l\le k}x_{k,l}^2+\left(\sum_{1\le l\le k}x_{k,l}\right)^2}\,\cdot\\ &\qquad\qquad\times\sqrt{\sum_k\sum_{1\le l\le k}\tilde{x}_{k,l}^2+\left(\sum_{1\le l\le k}\tilde{x}_{k,l}\right)^2}\\ &= \frac{s}{s+1}\|Ax\|\|A\tilde{x}\|,\end{aligned}$$

which shows that $\tau_s(x)\le s/(s+1)$ for every $x\in X_s$ and therefore $\tau_{s,s}\le s/(s+1)$.

In the following example, we construct an operator A that is bounded, injective, and has non-closed range in such a way that every set X_s is a set of well-posedness for A.

Example 4.2. *Let again $\Lambda := \{(k,l) : k,l\in\mathbb{N},\ 1\le l\le k\}$ and let*

$$\eta_k := \frac{1}{\sqrt{k}}\sum_{1\le l\le k}e^{k,l}.$$

Then the vectors η_k form an orthonormal system in $\ell^2(\Lambda)$. Choose now any sequence $\{c_k\}_{k\in\mathbb{N}}$ with $0<c_k<1$ for all k and $\lim_{k\to\infty}c_k=1$. Define $A\colon \ell^2(\Lambda)\to\ell^2(\Lambda)$ by

$$Ax = x-\sum_{k\in\mathbb{N}}c_k\langle x,\eta_k\rangle\eta_k.$$

Then we obtain, with the abbreviation $d_k := 2c_k-c_k^2$,

$$\begin{aligned}\|Ax\|^2 &= \|x\|^2-2\sum_{k\in\mathbb{N}}c_k\langle x,\eta_k\rangle^2+\sum_{k\in\mathbb{N}}c_k^2\langle x,\eta_k\rangle^2\\ &= \|x\|^2-\sum_{k\in\mathbb{N}}(2c_k-c_k^2)\langle x,\eta_k\rangle^2\\ &= \|x\|^2-\sum_{k\in\mathbb{N}}\left(\frac{d_k}{k}\left(\sum_{1\le l\le k}x_{k,l}\right)^2\right).\end{aligned}\tag{4.3}$$

In particular, $\|Ax\|^2\le\|x\|^2$, showing that A is bounded. Moreover, it is obvious that A is not boundedly invertible, as $\|A\eta_k\|^2 = 1-d_k\to 0$ as $k\to\infty$.

Now we show that every set X_s is a set of well-posedness for A. Because A is bounded, we have to show that $\sigma_s > 0$ for all $s \in \mathbb{N}$ and $\tau_s(x) < 1$ for all $s \in \mathbb{N}$ and $x \in X_s$. Let therefore $s \in \mathbb{N}$ and let $x \in X_s$. Because $|\mathrm{supp}(x)| \le s$, it follows that

$$\left(\sum_{1\le l\le k} x_{k,l}\right)^2 \le \min\{k,s\} \sum_{1\le l\le k} x_{k,l}^2.$$

Thus (4.3) *implies that*

$$\begin{aligned}\|Ax\|^2 &\ge \|x\|^2 - \sum_{k\in\mathbb{N}}\Big(\min\{k,s\}\frac{d_k}{k}\sum_{1\le l\le k} x_{k,l}^2\Big)\\ &= \sum_{k\in\mathbb{N}}\Big(\Big(1-\min\{k,s\}\frac{d_k}{k}\Big)\sum_{1\le l\le k} x_{k,l}^2\Big)\\ &\ge \inf_{k\in\mathbb{N}}\Big(1-d_k\frac{\min\{k,s\}}{k}\Big)\|x\|^2.\end{aligned}$$

Because the term d_k is strictly smaller than 1 and $\min\{k,s\}/k$ tends to zero as $k\to\infty$, it follows that

$$\sigma_s^2 \ge \inf_{k\in\mathbb{N}}\Big(1-d_k\frac{\min\{k,s\}}{k}\Big) > 0.$$

Now let $\tilde{x} \in X_s$ be such that $\mathrm{supp}(\tilde{x}) \cap \mathrm{supp}(x) = \emptyset$. Define the mapping $\pi_k \colon X \to X$ by

$$\pi_k(\hat{x}) := \sum_{1\le l\le k} \hat{x}_{k,l} e^{k,l}.$$

Then

$$\langle Ax, A\tilde{x}\rangle = \sum_{k\in\mathbb{N}} \langle A\pi_k x, A\pi_k \tilde{x}\rangle.$$

Moreover,

$$\begin{aligned}\langle A\pi_k x, A\pi_k\tilde{x}\rangle &= \langle \pi_k x, \pi_k \tilde{x}\rangle - d_k\langle x,\eta_k\rangle\langle\tilde{x},\eta_k\rangle\\ &= -d_k\langle x,\eta_k\rangle\langle\tilde{x},\eta_k\rangle = -d_k\left(\sum_{1\le l\le k} x_{k,l}\right)\left(\sum_{1\le l\le k}\tilde{x}_{k,l}\right).\end{aligned} \tag{4.4}$$

Now denote

$$n_k := |\mathrm{supp}(x)\cap\mathrm{supp}(\eta_k)| \qquad \textit{and} \qquad \tilde{n}_k := |\mathrm{supp}(\tilde{x})\cap\mathrm{supp}(\eta_k)|.$$

Then (4.4) *implies that*

$$\langle A\pi_k x, A\pi_k\tilde{x}\rangle^2 \le d_k^2\frac{n_k\tilde{n}_k}{k^2}\left(\sum_{1\le l\le k} x_{k,l}^2\right)\left(\sum_{1\le l\le k}\tilde{x}_{k,l}^2\right). \tag{4.5}$$

Moreover, as $d_k n_k/k<1$ *and* $d_k\tilde n_k/k<1$ *for all* k,

$$\|A\pi_k x\|^2\|A\pi_k\tilde x\|^2 \ge \Big(1-d_k\frac{n_k}{k}\Big)\Big(1-d_k\frac{\tilde n_k}{k}\Big)\left(\sum_{1\le l\le k} x_{k,l}^2\right)\left(\sum_{1\le l\le k}\tilde x_{k,l}^2\right). \tag{4.6}$$

Now define

$$\theta_s := \sup_{k\in\mathbb{N}} d_k\frac{\min\{s,k/2\}}{k}.$$

Then

$$\theta_s = \max\Big\{\max_{1\le k\le 2s}\frac{d_k}{2},\sup_{k>2s}\frac{d_k s}{k}\Big\} \le \max\Big\{\max_{1\le k\le 2s}\frac{d_k}{2},\frac{s}{2s+1}\Big\} < \frac{1}{2},$$

as $0<d_k<1$ *for all* $k\in\mathbb{N}$. *Next, note that the assumptions* $|\mathrm{supp}(x)|\le s$, $|\mathrm{supp}(\tilde x)|\le s$, *and* $\mathrm{supp}(x)\cap\mathrm{supp}(\tilde x)=\emptyset$ *imply that* $n_k\tilde n_k\le\min\{s^2,k^2/4\}$ *for all* $k\in\mathbb{N}$. *Thus we obtain that for all* $k\in\mathbb{N}$

$$d_k^2\frac{n_k\tilde n_k}{k^2}\le\theta_s^2<\frac14.$$

Now, the inequalities $0\le d_k n_k/k<1$, $0\le d_k\tilde n_k/k<1$, *and* $d_k^2 n_k\tilde n_k/k^2\le\theta$ *imply that for all* $k\in\mathbb{N}$

$$\Big(1-d_k\frac{n_k}{k}\Big)\Big(1-d_k\frac{\tilde n_k}{k}\Big)\ge(1-\theta_s)^2.$$

Consequently, we obtain from (4.5) *and* (4.6) *that*

$$\langle A\pi_k x, A\pi_k\tilde x\rangle^2 \le \frac{\theta_s^2}{(1-\theta_s)^2}\|Ax\|^2\|A\tilde x\|^2.$$

Consequently,

$$\begin{aligned}\langle Ax, A\tilde x\rangle &= \sum_{k\in\mathbb{N}}\langle A\pi_k x, A\pi_k\tilde x\rangle\\ &\le \frac{\theta_s}{1-\theta_s}\sum_{k\in\mathbb{N}}\|A\pi_k x\|\|A\pi_k\tilde x\|\\ &\le \frac{\theta_s}{1-\theta_s}\sqrt{\sum_{k\in\mathbb{N}}\|A\pi_k x\|^2}\sqrt{\sum_{k\in\mathbb{N}}\|A\pi_k\tilde x\|^2}\\ &= \frac{\theta_s}{1-\theta_s}\|Ax\|\|A\tilde x\|.\end{aligned}$$

Because $\theta_s<1/2$, *the assertion follows with*

$$\tau_{s,s}\le\frac{\theta_s}{1-\theta_s}<1.$$

5. Conclusion

We have derived a characterisation of those linear operators between ℓ^2 spaces and general Hilbert spaces for which the assumption of sparsity constraints leads to well-posed problems. For this well-posedness to hold, the operator A as well as its inverse have to be bounded on the set X_s consisting of all s-sparse elements of $\ell^2(\Lambda)$. In addition, if x, $\tilde{x} \in X_s$ have disjoint support, then the angle between Ax and $A\tilde{x}$ has to be strictly larger than some positive number. These conditions are closely related to various formulations of a restricted isometry property that is commonly encountered in the context of compressed sensing. There, this property implies first the stability of sparse regularization and, second, that its solution can be computed by minimising the ℓ^1-norm instead. The results of the present paper indicate that conditions that are similar to a restricted isometry property appear naturally when treating sparse regularization problems. Also, the examples show that, although the approximate inversion of compact operators with sparse regularization can lead to problems, the theory is not restricted to invertible operators. Instead, there exist operators with non-closed range for which every set X_s is a well-posedness class, that is, any restriction of the support of the solution yields a well-posed problem.

Acknowledgement. This work has been supported by the Austrian Science Fund (FWF) within the national research network *Industrial Geometry*, project 9203-N12. The author wants to thank the referees for their valuable remarks, in particular concerning relations between the quantities used in the statement of the main theorem and similar quantities appearing in the theory of compressed sensing.

REFERENCES

[1] K. Bredies and D. Lorenz, *Regularization with non-convex separable constraints*, Inv. Prob., 25(8), 085011 (14pp), 2009.

[2] E.J. Candès, *Compressive sampling*, in International Congress of Mathematicians, Vol. III, Eur. Math. Soc., Zürich, 1433–1452, 2006.

[3] E.J. Candès, *The restricted isometry property and its implications for compressed sensing*, C.R. Math. Acad. Sci. Paris, 346(9-10), 589–592, 2008.

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

[5] E.J. Candès and T. Tao, *Decoding by linear programming*, IEEE Trans. Inf. Theory, 51(12), 4203–4215, 2005.

[6] S.S. Chen, D.L. Donoho, and M.A. Saunders, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput., 20(1), 33–61 (electronic), 1998.

[7] I. Daubechies, M. Defrise, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Commun. Pure Appl. Math., 57(11), 1413–1457, 2004.

[8] 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.

[9] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, New York, 2010.

[10] M. Grasmair, *Well-posedness and convergence rates for sparse regularization with sublinear l^q penalty term*, Inv. Prob. Imag., 3(3), 383–387, 2009.

[11] M. Grasmair, *Non-convex sparse regularization*, J. Math. Anal. Appl., 365(1), 19–28, 2010.

[12] M. Grasmair, M. Haltmeier, and O. Scherzer, *Sparse regularization with l^q penalty term*, Inv. Prob., 24(5), 055020, 13, 2008.

[13] 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, 2009.

[14] R. Gribonval and M. Nielsen, *Beyond sparsity: Recovering structured representations by l^1 minimization and greedy algorithms*, Adv. Comput. Math., 28(1), 23–41, 2008.

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

[16] D.A. Lorenz, S. Schiffler, and D. Trede, *Beyond convergence rates: Exact inversion with Tikhonov regularization with sparsity constraints*, Inv. Prob., to appear.

[17] E. Ramlau and R. Resmerita, *Convergence rates for regularization with sparsity constraints*, Electron. Trans. Numer. Anal., 37, 87–104, 2010.

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

[19] A.N. Tikhonov and V.Y. Arsenin, *Solutions of Ill-Posed Problems*, John Wiley & Sons, Washington, D.C., 1977.

[20] J.A. Tropp, *Greed is good: Algorithmic results for sparse approximation*, IEEE Trans. Inf. Theory, 50(10), 2231–2242, 2004.

[21] C.A. Zarzer, *On Tikhonov regularization with non-convex sparsity constraints*, Inv. Prob., 25, 025006, 2009.