Linear Convergence Rates for Tikhonov Regularisation with Positively Homogeneous Functionals

Markus Grasmair

Computational Science Center
University of Vienna
Nordbergstr. 15
A–1090 Vienna, Austria

May 4, 2011

Abstract

The goal of this paper is the formulation of an abstract setting that can be used for the derivation of linear convergence rates for a large class of sparsity promoting regularisation functionals for the solution of ill-posed linear operator equations. Examples where the proposed setting applies include joint sparsity and group sparsity, but also (possibly higher order) discrete total variation regularisation. In all these cases, a range condition together with some kind of restricted injectivity imply linear convergence rates. That is, the error in the approximate solution, measured in terms of the regularisation functional, is of the same order as the noise level.

1 Introduction

Recently, sparsity promoting regularisation methods for the solution of ill-posed linear operator equations of the form

$$Ax = y \tag{1}$$

have become popular (see for instance [12, 17]). There one assumes that the true solution $x^\dagger$ of (1) is sparse with respect to some basis or frame $(e_\lambda)_{\lambda\in\Lambda}$ of the domain of the operator A to be inverted. In order to obtain a stable, sparse approximation of $x^\dagger$ in presence of noisy data y^δ satisfying $\|y - y^\delta\| \le \delta$, one may then minimise the Tikhonov functional

$$\mathcal{T}_{\alpha,y^\delta}(x) := \|Ax - y\|^2 + \alpha \sum_{\lambda\in\Lambda} |\langle e_\lambda, x\rangle| \tag{2}$$

for some suitable regularisation parameter $\alpha > 0$ depending on δ.

One interesting feature of the sparsity promoting regularisation method (2) is that, under some additional assumptions, the accuracy of the regularised solutions x_α^δ, that is, the minimisers of $\mathcal{T}_{\alpha,y^\delta}$, is of the same order as the noise level δ (see [20, 22, 23]). Similarly, there are results in the context of compressed sensing that yield conditions that guarantee that the residual method

$$\sum_{\lambda \in \Lambda} |\langle e_\lambda, x\rangle| \to \min \qquad\qquad \text{subject to } \|Ax - y^\delta\| \le \delta$$

implies linear estimates $\|x_\alpha^\delta - x^\dagger\| = O(\delta)$ (see [6, 7, 10, 13, 14, 21, 23]). In constrast, for quadratic Tikhonov regularisation, the best possible rate is $\|x_\alpha^\delta - x^\dagger\| = O(\delta^{2/3})$ unless the operator A has finite rank (see [28]).

This paper extends the results on linear convergence rates for sparsity promoting regularisation to more general situations that are not covered by the existing theory. One generalisation is concerned with group sparsity. There one assumes that the indices λ can be assigned to different groups, and one wants the number of groups containing non-zero coefficients to be small, while no sparsity is required within the groups. As another example, discrete total variation regularisation is studied, which can be interpreted as regularisation promoting sparsity of the discrete gradient. Also there, linear convergence rates are possible under suitable assumptions.

In Section 2 we prove a result on linear convergence rates in a quite abstract setting (see Theorem 2.6). From Section 3 onwards, we consider a setting that can be directly applied to group sparsity and discrete total variation regularisation. First, we prove the well-posedness of sparse regularisation in this setting, which is especially important for discrete total variation regularisation, where the regularisation functional itself is not coercive, and thus the existence of regularised solution can only be guaranteed by suitable assumptions on the operator A. Having assured the well-posedness of the regularisation method, we then apply the convergence rates result of Section 2 (see Theorem 4.4). Finally, we discuss in Sections 5–7, the application to group sparsity, joint sparsity and discrete total variation regularisation.

2 Linear Convergence Rates

In this article we consider the solution of an ill-posed linear operator equation

$$Ax = y\,,$$

where $A\colon X \to Y$ is a bounded linear operator between two Banach spaces X and Y, by means of Tikhonov regularisation with a convex and positively homogeneous regularisation term $\mathcal{R}\colon X \to [0, +\infty]$. That is, we solve the equation $Ax = y$ approximately by minimising, for some regularisation parameter $\alpha > 0$, the Tikhonov functional

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

We assume in the following that $y \in Y$ is fixed and denote by $x^\dagger$ any $\mathcal{R}$-minimising solution of the equation $Ax = y$, that is,

$$x^\dagger \in \arg\min\bigl\{\mathcal{R}(x) : Ax = y\bigr\}\,.$$

In addition, we denote by x_α^δ any regularised solution of the equation $Ax = y^\delta$ with regularisation parameter $\alpha > 0$, where the right hand side is allowed to differ from the true data y by at most $\delta > 0$. More precisely, we assume that $x_\alpha^\delta \in X$ is any element satisfying

$$x_\alpha^\delta \in \arg\min\big\{\mathcal{T}_{\alpha,y^\delta}(x) : x \in X\big\} \qquad \text{for some } y^\delta \in Y \text{ with } \|y^\delta - y\| \leq \delta\ .$$

The question of convergence rates is then concerned with quantifying the difference between x_α^δ and $x^\dagger$ for small noise level δ and a regularisation parameter α adapted to δ.

Assumption 2.1. We assume that the following conditions are satisfied:

1. The Tikhonov functional $\mathcal{T}_{\alpha,\tilde{y}}$ is coercive for every $\alpha > 0$ and $\tilde{y} \in Y$ in the sense that its lower level sets are weakly pre-compact in X.
2. The spaces X and Y are Banach spaces and the mapping $A\colon X \to Y$ is bounded linear.
3. The functional $\mathcal{R}\colon X \to [0, +\infty]$ is convex, lower semicontinuous, and positively homogeneous, that is, $\mathcal{R}(0) = 0$ and $\mathcal{R}(tx) = t\mathcal{R}(x)$ for every $x \in X$ and $t > 0$. ■

Remark 2.2. In what follows, the coercivity of $\mathcal{T}_{\alpha,\tilde{y}}$ is never needed explicitly, but it is essential nevertheless. It implies that the functional $\mathcal{T}_{\alpha,\tilde{y}}$ attains its minimum and therefore Tikhonov regularisation is well-defined: The convexity and lower semi-continuity of $\mathcal{R}$ imply that $\mathcal{R}$ is weakly lower semi-continuous. Also, the mapping $x \mapsto \|Ax - \tilde{y}\|^2$ is weakly lower semi-continuous, because A is bounded. Thus $\mathcal{T}_{\alpha,\tilde{y}}$ is weakly lower semi-continuous. Therefore, the coercivity of $\mathcal{T}_{\alpha,\tilde{y}}$ allows the application of the direct method in the calculus of variations (see for instance [11, Chpt. 1]), which proves the existence of a minimiser. ■

Starting with the pioneering paper by Burger and Osher [5], convergence rates for convex regularisation methods on Banach spaces have typically been derived with respect to the *Bregman distance*, which measures the distance between $\mathcal{R}$ and an affine approximation to $\mathcal{R}$ at the true solution $x^\dagger$. It is defined, for $\xi \in \partial\mathcal{R}(x^\dagger)$, as

$$D^\xi(\tilde{x}; x^\dagger) := \mathcal{R}(\tilde{x}) - \mathcal{R}(x^\dagger) - \langle \xi, \tilde{x} - x^\dagger \rangle\ .$$

In the case of positively homogeneous regularisation terms, the Bregman distance assumes a very simple form:

Lemma 2.3. *Let $x \in X$ and $\xi \in X^*$. Then $\xi \in \partial\mathcal{R}(x)$, if and only if $\xi \in \partial\mathcal{R}(0)$ and $\mathcal{R}(x) = \langle \xi, x \rangle$. In particular,*

$$D^\xi(\tilde{x}; x) = \mathcal{R}(\tilde{x}) - \langle \xi, \tilde{x} \rangle$$

for every $\tilde{x} \in X$.

Proof. Assume that $\xi \in \partial\mathcal{R}(x)$. Then

$$\mathcal{R}(\tilde{x}) \geq \mathcal{R}(x) + \langle \xi, \tilde{x} - x \rangle \tag{3}$$

for every $\tilde{x} \in X$. Choosing $\tilde{x} = 0$, we obtain $\mathcal{R}(x) \leq \langle \xi, x \rangle$. Conversely, if $\tilde{x} = 2x$, the positive homogeneity of $\mathcal{R}$ implies the inequality $\mathcal{R}(x) \geq \langle \xi, x \rangle$, proving that, in fact, equality holds. Thus (3) simplifies to the inequality $\mathcal{R}(\tilde{x}) \geq \langle \xi, \tilde{x} \rangle$ for every $\tilde{x} \in X$, which implies that $\xi \in \partial\mathcal{R}(0)$.

Now assume that $\xi \in \partial\mathcal{R}(0)$ and $\mathcal{R}(x) = \langle \xi, x \rangle$. Then

$$\mathcal{R}(\tilde{x}) \geq \langle \xi, \tilde{x} \rangle = \langle \xi, \tilde{x} \rangle + \mathcal{R}(x) - \langle \xi, x \rangle = \mathcal{R}(x) + \langle \xi, \tilde{x} - x \rangle$$

for every $\tilde{x} \in X$, which shows that $\xi \in \partial\mathcal{R}(x)$. □

Using this result, the convergence rates in terms of the Bregman distance can be written as follows:

Lemma 2.4. *Assume that $x^\dagger$ satisfies a range condition with $\xi \in \operatorname{Ran} A^* \cap \partial\mathcal{R}(x^\dagger)$. Then we have for a parameter choice $\alpha \sim \delta$ that*

$$\mathcal{R}(x_\alpha^\delta) - \langle \xi, x_\alpha^\delta \rangle = O(\delta) \qquad \textit{and} \qquad \|A(x_\alpha^\delta - x^\dagger)\| = O(\delta)$$

as $\delta \to 0$.

Proof. The assertion follows from Lemma 2.3 and standard results on regularisation in Banach spaces (see [24, 30]). □

We will now use Lemma 2.4 for deriving convergence rates not with respect to the Bregman distance but with respect to the regularisation functional $\mathcal{R}$ itself. Denote to that end for $\xi \in X^*$ the set $K_\xi \subset X$ defined by

$$K_\xi := \bigl\{x \in X : \xi \in \partial\mathcal{R}(x)\bigr\} .$$

Lemma 2.5. *If $\xi \notin \partial\mathcal{R}(0)$, then $K_\xi = \emptyset$. Conversely, if $\xi \in \partial\mathcal{R}(0)$, then*

$$x \in K_\xi \qquad \Longleftrightarrow \qquad \langle \xi, x \rangle = \mathcal{R}(x) . \tag{4}$$

In this case, K_ξ is a non-empty convex and closed cone.

Proof. Lemma 2.3 implies that $\partial\mathcal{R}(x) \subset \partial\mathcal{R}(0)$ for every $x \in X$, which in turn implies that $K_\xi = \emptyset$ whenever $\xi \notin \partial\mathcal{R}(0)$. Conversely, if $\xi \in \partial\mathcal{R}(0)$, then we obtain from Lemma 2.3 that $\xi \in \partial\mathcal{R}(x)$ if and only if $\mathcal{R}(x) = \langle \xi, x \rangle$, proving (4).

Now assume that $\xi \in \partial\mathcal{R}(0)$ (which implies that $0 \in K_\xi$). Then in particular $\mathcal{R}(x) \geq \langle \xi, x \rangle$ for every $x \in X$. Thus it is also possible to write

$$K_\xi = \bigl\{x \in X : \mathcal{R}(x) - \langle \xi, x \rangle \leq 0\bigr\} .$$

Because $\mathcal{R}$ is lower semi-continuous, it follows that K_ξ is a closed set. Moreover, the positive homogeneity of $\mathcal{R}$ and (4) directly show that K_ξ is a cone. In order to show that K_ξ is convex, let x, $\tilde{x} \in K_\xi$ and $0 < \lambda < 1$. Then

$$\begin{aligned} \mathcal{R}\bigl(\lambda x + (1-\lambda)\tilde{x}\bigr) &\geq \bigl\langle \xi, \lambda x + (1-\lambda)\tilde{x} \bigr\rangle = \mathcal{R}(\lambda x) + \mathcal{R}\bigl((1-\lambda)\tilde{x}\bigr) \\ &= \frac{1}{2}\bigl(\mathcal{R}(2\lambda x) + \mathcal{R}(2(1-\lambda)\tilde{x})\bigr) \geq \mathcal{R}\bigl(\lambda x + (1-\lambda)\tilde{x}\bigr) , \end{aligned}$$

which shows that

$$\mathcal{R}\bigl(\lambda x + (1-\lambda)\tilde{x}\bigr) = \bigl\langle \xi, \lambda x + (1-\lambda)\tilde{x} \bigr\rangle$$

and therefore $\lambda x + (1-\lambda)\tilde{x} \in K_\xi$, proving the assertion. □

In particular, this last result shows that for $\xi \in \mathcal{R}(x^\dagger)$ we have

$$K_\xi = \{x \in X : D^\xi(x; x^\dagger) = 0\} .$$

In other words, the cone K_ξ consists precisely of those elements in X that cannot be distinguished by means of the Bregman distance D^ξ. Thus, using the preceding result, one sees that all the information about the behaviour of possible regularised solutions in K_ξ as the noise level δ and the regularisation parameter α tend to zero is contained in the second term $\|A(x_\alpha^\delta - x^\dagger)\|$. In order to derive stronger convergence rates, it is therefore necessary to introduce additional postulates concerning the separation properties of the operator A on the set K_ξ. In the results on linear convergence rates for sparse regularisation derived in [20, 23], this separation was obtained by requiring a *restricted injectivity* condition to hold on the support of the element $x^\dagger$. The following result replaces the set where injectivity of A should hold by the (possibly enlarged) cone K_ξ.

In addition, we need that the regularisation term $\mathcal{R}$ grows sufficiently fast away from K_ξ. This can be formulated as a condition on the behaviour of the sub-differential $\mathcal{R}(0)$ on the normal cone of K_ξ. Here we define the *normal cone* $\mathcal{N}(K)$ of a set $K \subset X$ as

$$\mathcal{N}(K) := \{\zeta \in X^* : \langle \zeta, x\rangle \le 0 \text{ for all } x \in K\} \subset X^* .$$

Similarly, if $\Xi \subset X^*$, we define

$$\mathcal{N}(\Xi) := \{x \in X : \langle \zeta, x\rangle \le 0 \text{ for all } \zeta \in \Xi\} \subset X .$$

Theorem 2.6. *Assume that $x^\dagger$ satisfies the range condition with $\xi \in \operatorname{Ran} A^* \cap \partial\mathcal{R}(x^\dagger)$. Assume moreover that there exists a locally compact, closed and convex cone $\tilde{K}$ and some $C_0 > 0$ such that the following hold:*

$$K_\xi \subset \tilde{K} \subset x^\dagger + \operatorname{Dom}(\mathcal{R}) , \tag{5}$$

$$\|A(\tilde{x} - x^\dagger)\| \ge C_0 \|\tilde{x} - x^\dagger\| \qquad \textit{for all } \tilde{x} \in \tilde{K} , \tag{6}$$

$$c_0 := \sup\{c \ge 0 : (\xi + c\,\partial\mathcal{R}(0) \cap \mathcal{N}(\tilde{K})) \subset \partial\mathcal{R}(0)\} > 0 . \tag{7}$$

Then we have for a parameter choice $\alpha \sim \delta$ that

$$\mathcal{R}(x_\alpha^\delta - x^\dagger) = O(\delta) \qquad \textit{as } \delta \to 0 .$$

Proof. We show in the following that there exists constants C_1, $C_2 > 0$ such that

$$\mathcal{R}(x - x^\dagger) \le C_1 \|A(x - x^\dagger)\| + C_2\big(\mathcal{R}(x) - \langle \xi, x\rangle\big) \tag{8}$$

for every $x \in X$. Then the assertion will follow from Lemma 2.4.

Note first that the homogeneity of $\mathcal{R}$ implies its sub-additivity, and therefore

$$\mathcal{R}(x - x^\dagger) \le \inf\{\mathcal{R}(x - \tilde{x}) + \mathcal{R}(\tilde{x} - x^\dagger) : \tilde{x} \in \tilde{K}\} .$$

Now note that (6) together with the facts that $\tilde{K}$ is locally compact and $\tilde{K} \subset x^\dagger + \operatorname{Dom}(\mathcal{R})$ implies that there exists a constant $C_1 > 0$ such that

$$\mathcal{R}(\tilde{x} - x^\dagger) \le C_1 \|A(\tilde{x} - x^\dagger)\| \qquad \text{for all } \tilde{x} \in \tilde{K} .$$

Moreover, the coercivity and positive homogeneity of $\mathcal{R}$ imply the existence of $\tilde{C}^{(1)} > 0$ such that

$$\|A(\tilde{x}-x^\dagger)\| \le \|A(x-x^\dagger)\| + \|A\|\|\tilde{x}-x\| \le \|A(x-x^\dagger)\| + \tilde{C}^{(1)}\|A\|\mathcal{R}(x-\tilde{x}) \ .$$

Thus, with $\tilde{C}^{(2)} := (1 + C_1\tilde{C}^{(1)}\|A\|)$ it follows that

$$\mathcal{R}(x-x^\dagger) \le C_1\|A(x-x^\dagger)\| + \tilde{C}^{(2)} \inf\bigl\{\mathcal{R}(x-\tilde{x}) : \tilde{x}\in\tilde{K}\bigr\} \ .$$

Now we apply standard results from convex analysis to see that we have for every $c > 0$ that

$$\begin{aligned}
&\inf\bigl\{\mathcal{R}(x-\tilde{x}) : \tilde{x}\in\tilde{K}\bigr\} \\
&\qquad = -\inf\bigl\{-\langle\eta,x\rangle + \mathcal{R}^*(\eta) : \eta\in\mathcal{N}(\tilde{K})\bigr\} \\
&\qquad = -\inf\bigl\{-\langle\eta,x\rangle : \eta\in\mathcal{N}(\tilde{K})\cap\partial\mathcal{R}(0)\bigr\} \\
&\qquad = \sup\bigl\{\langle\eta,x\rangle : \eta\in\mathcal{N}(\tilde{K})\cap\partial\mathcal{R}(0)\bigr\} \\
&\qquad = \sup\bigl\{\langle c\xi+\eta,x\rangle : \eta\in\mathcal{N}(\tilde{K})\cap\partial\mathcal{R}(0)\bigr\} - c\langle\xi,x\rangle \\
&\qquad = c\sup\bigl\{\langle\xi+\eta,x\rangle : \eta\in c\partial\mathcal{R}(0)\cap\mathcal{N}(\tilde{K})\bigr\} - c\langle\xi,x\rangle \ .
\end{aligned} \tag{9}$$

Now note that, for $\xi+\eta\in\partial\mathcal{R}(0)$ we have that $\langle\xi+\eta,x\rangle \le \mathcal{R}(x)$. Because the definition of c_0 implies that $\xi + (c_0/2)\partial\mathcal{R}(0)\cap\mathcal{N}(\tilde{K}) \subset \partial\mathcal{R}(0)$, we obtain from (9) that

$$\inf\bigl\{\mathcal{R}(x-\tilde{x}) : \tilde{x}\in\tilde{K}\bigr\} \le \frac{c_0}{2}\bigl(\mathcal{R}(x) - \langle\xi,x\rangle\bigr) \ .$$

This proves the assertion with $C_2 = \tilde{C}^{(2)}c_0/2$. □

Remark 2.7. Note that (6) holds if $\tilde{K}$ is finite dimensional and the restriction of A to the subspace spanned by $\tilde{K}$ is injective. This is precisely the situation that will be considered in Section 4 below. ■

Remark 2.8. In the setting of standard ℓ^1-regularisation, we have $X = \ell^2(\Lambda)$ for some countable index set Λ and $\mathcal{R}(x) = \sum_\lambda |x_\lambda|$. Then we have $\xi\in\partial\mathcal{R}(x^\dagger) \subset \ell^2(\Lambda)$, if and only if $\xi_\lambda = \operatorname{sgn}(x_\lambda)$ for $x_\lambda \neq 0$, $\xi_\lambda \in [-1,1]$ for $x_\lambda = 0$, and $\sum_\lambda \xi_\lambda^2 < +\infty$. Moreover,

$$K_\xi = \bigl\{x\in\ell^2(\Lambda) : x_\lambda \ge 0 \text{ if } \xi_\lambda = 1,\ x_\lambda \le 0 \text{ if } \xi_\lambda = -1, \text{ and } x_\lambda = 0 \text{ else}\bigr\} ,$$

which is a locally compact cone, because the condition $\sum_\lambda \xi_\lambda^2 < +\infty$ implies that $\Omega := \bigl\{\lambda\in\Lambda : |\xi_\lambda| < 1\bigr\}$ is a finite set. Now let $\tilde{K}$ be the finite dimensional vector space $\tilde{K} := \operatorname{span} K_\xi \simeq \ell^2(\Omega)$. Then $\mathcal{N}(\tilde{K}) \simeq \ell^2(\Lambda\setminus\Omega)$. Because $\partial\mathcal{R}(0) = \bigl\{\zeta\in X : |\zeta_\lambda| \le 1 \text{ for all } \lambda\in\Lambda\bigr\}$, it follows that c_0 defined in (7) satisfies

$$c_0 = \sup\bigl\{c \ge 0 : |\xi_\lambda| + c \le 1 \text{ for all } \lambda\in\Lambda\setminus\Omega\bigr\} = 1 - \max\bigl\{|\xi_\lambda| : \lambda\in\Lambda\setminus\Omega\bigr\} > 0 \ .$$

In particular, the condition (7) naturally holds for standard ℓ^1-regularisation. Also note that the same term $1 - \max\bigl\{|\xi_\lambda| : \lambda\in\Lambda\setminus\Omega\bigr\}$ appears in the proof for the linear convergence of ℓ^1-regularisation (see [20, Proof of Thm. 15]). ■

3 Well-posedness

From now on we consider the setting where the regularisation term $\mathcal{R}$ is the composition of a linear mapping L and a regularisation functional on the range of L. More precisely, we assume that $L\colon D \subset X \to Z$ is a linear mapping between a subspace D of X and a Banach space Z, that $\mathcal{S}\colon Z \to [0,+\infty]$ is some mapping, and

$$\mathcal{R}(x) = \begin{cases} \mathcal{S}(Lx) & \text{if } x \in D\,, \\ +\infty & \text{else.} \end{cases} \tag{10}$$

Before applying the convergence rates result Theorem 2.6, we first have to assert that the conditions of Assumption 2.1 hold. A particular problem is the coercivity of the functional $\mathcal{T}_{\alpha,\tilde{y}}$, which will necessarily fail if the mapping L is not injective, unless the operator A behaves well on the kernel of L.

For instance, in the case of l-th order total variation regularisation on a bounded domain Ω, the kernel of L consists precisely of the polynomials of degree smaller than l. The coercivity of $\mathcal{T}_{\alpha,\tilde{y}}$ then requires that the operator A does not annihilate these polynomials (see [1, 2, 9] for first order and [30, Prop. 3.70] for higher order regularisation). An analogous condition can also be used in the abstract setting, as the following results show. Very similar results have also been derived in [4].

Also, the coercivity of $\mathcal{T}_{\alpha,\tilde{y}}$ almost immediately implies that Tikhonov regularisation with the regularisation term $\mathcal{R}$ is a well-posed regularisation method in the sense of [16, 30, 33]. That is, the functional $\mathcal{T}_{\alpha,\tilde{y}}$ has a minimiser that depends, for fixed parameter $\alpha > 0$, continuously on the data $\tilde{y}$, and converges to a true solution when the error in the right hand side $\tilde{y}$ and the regularisation parameter α both converge to zero in a suitable manner.

Assumption 3.1. We assume that the following conditions are satisfied:

1. The functional $\mathcal{R}$ has the form described in (10).
2. The Banach space X is reflexive.
3. There exists a closed subspace $W \subset X$ such that $X = W \oplus \operatorname{Ker} L$ and the restriction of $\mathcal{R}$ to the Banach space W is coercive.
4. The mapping $A\colon X \to Y$ is bounded linear.
5. The restriction of A to $\operatorname{Ker} L$ is boundedly invertible.
6. The mapping $\mathcal{R}$ is weakly lower semi-continuous. ■

Note that we waive for the moment the assumption that the operator $\mathcal{R}$ (or $\mathcal{S}$) is positively homogeneous.

Proposition 3.2. *Let Assumption 3.1 hold. Then, for every $\alpha > 0$ and $y \in Y$, the functional $\mathcal{T}_{\alpha,y}$ is coercive.*

Proof. Note first that the reflexivity of X implies that the mapping $\mathcal{T}_{\alpha,y}$ is coercive, if and only if its lower level sets are bounded in X.

Since the restriction of A to $\operatorname{Ker} L$ is invertible, there exists $C_1 > 0$ such that $C_1\|Ax\| \geq \|x\|$ for all $x \in \operatorname{Ker} L$. Now let $P\colon X \to X$ be the projection onto $\operatorname{Ker} L$ along W. That is, if $x \in X$ is decomposed as $x = x_L + x_W$ with $x_L \in \operatorname{Ker} L$ and $x_W \in W$, then $Px = x_L$. Denote moreover for simplicity $P' := \operatorname{Id} - P$.

Let now $\alpha > 0$, $y \in Y$, and $t > 0$ be fixed. Because the restriction of $\mathcal{R}$ to W is coercive, and $\mathcal{R}(x) = \mathcal{R}(P'x)$ for every $x \in X$, there exists $C_2 > 0$ such that

$$\|P'x\| \le C_2 \qquad \text{whenever } x \in X \text{ satisfies } \alpha\mathcal{R}(x) \le t\,.$$

Assume that $x \in X$ satisfies $\mathcal{T}_{\alpha,y}(x) \le t$. Then in particular $\alpha\mathcal{R}(x) \le t$ and consequently $\|P'x\| \le C_2$. Thus

$$\begin{aligned}
\|x\| &= \|Px + P'x\| \\
&\le \|Px\| + \|P'x\| \\
&\le C_1\|APx\| + C_2 \\
&= C_1\|Ax - AP'x\| + C_2 \\
&\le C_1\big(\|Ax - y + y\| + \|A\|\|P'x\|\big) + C_2 \\
&\le C_1\big(\sqrt{t} + \|y\| + C_2\|A\|\big) + C_2\,.
\end{aligned}$$

This proves the coercivity of $\mathcal{T}_{\alpha,y}$. □

Remark 3.3. The assumption that $\operatorname{Ker} L$ is closed is satisfied as soon as the mapping L has a closed graph. Moreover, the decomposition $X = W \oplus \operatorname{Ker} L$ is always possible if X is a Hilbert space or if $\operatorname{Ker} L$ is finite dimensional. In the latter case, the assumption that the restriction of A to $\operatorname{Ker} L$ is invertible reduces to the assumption that $\operatorname{Ker} L \cap \operatorname{Ker} A = \{0\}$. ■

Remark 3.4. Assume that X and Z are reflexive, that L has a closed graph, and that the mapping $\mathcal{S}\colon Z \to [0, +\infty]$ is weakly lower semi-continuous and coercive. Then also the mapping $\mathcal{R}$ is weakly lower semi-continuous (see for instance [21, Prop. 3.3]). ■

Remark 3.5. Assume that L has a closed graph, and that $\operatorname{Ran} L$ is closed. Then it follows that $\operatorname{Ran} L$ and $X/\operatorname{Ker} L$ are isomorphic Banach spaces. Now assume that in addition $\operatorname{Ran} L$ is complemented in X by the closed linear space W. Then it follows that W and $X/\operatorname{Ker} L$, and therefore also W and $\operatorname{Ran} L$ are isomorphic. Consequently, under this assumptions the restriction of $\mathcal{R}$ to W is coercive, if and only if the restriction of $\mathcal{S}$ to $\operatorname{Ran} L$ is coercive. In particular, the coercivity of $\mathcal{R}$ follows from the coercivity of $\mathcal{S}$ (on the whole space Z). ■

For the next result recall that the functional $\mathcal{T}_{\alpha,y}$ satisfies the *Radon–Riesz property*, if the weak convergence of a sequence $(x_k)_{k\in\mathbb{N}} \subset X$ to $x \in X$ together with the convergence $\mathcal{T}_{\alpha,y}(x_k) \to \mathcal{T}_{\alpha,y}(x)$ implies that $(x_k)_{k\in\mathbb{N}}$ converges *strongly* to x (see [27]).

Proposition 3.6. *Let Assumption 3.1 hold. In addition, assume that $\mathcal{T}_{\alpha,y}$ satisfies the Radon–Riesz property. Then Tikhonov regularisation with $\mathcal{R}$ is stable and convergent.*

Proof. This is along the lines of the proofs of Theorems 3.23 and 3.26 in [30]. □

In the next result, it is shown that the Radon–Riesz property of $\mathcal{T}_{\alpha,y}$ holds, if Y is sufficiently regular and the functional $\mathcal{S}$ has the Radon–Riesz property. While not applicable to continuous total variation regularisation, the result can be used for proving the stability and convergence of *discrete* gradient based regularisation methods.

Lemma 3.7. *Let Assumption 3.1 hold. In addition, assume that Y is a uniformly convex Banach space, that Z is reflexive, that $\operatorname{Ran} L$ is closed and that $\mathcal{S}$ is coercive and satisfies the Radon–Riesz property. Then the functional $\mathcal{T}_{\alpha,y}$ satisfies the Radon–Riesz property.*

Proof. Let $(x_k)_{k\in\mathbb{N}} \subset X$ be any sequence weakly converging to some $x \in X$ such that $\mathcal{T}_{\alpha,y}(x_k) \to \mathcal{T}_{\alpha,y}(x) < \infty$. Because the mappings $\mathcal{R}$ and $x \mapsto \|Ax - y\|^2$ are weakly lower semi-continuous, it follows that also $\mathcal{R}(x_k) \to \mathcal{R}(x) < \infty$ and $\|Ax_k - y\|^2 \to \|Ax - y\|^2$.

Because the mapping A is bounded linear, it maps weakly convergent sequences to weakly convergent sequences, which implies that $Ax_k \rightharpoonup Ax$. The uniform convexity of Y implies that the norm on Y satisfies the Radon–Riesz property (see [27, Thm. 5.2.18]). Thus the assumptions $Ax_k \rightharpoonup Ax$ and $\|Ax_k - y\| \to \|Ax - y\|$ together imply that $Ax_k \to Ax$.

Since $\mathcal{S}$ is coercive and Z is reflexive, it follows that the sequence $(Lx_k)_{k\in\mathbb{N}} \subset Z$ is weakly pre-compact. Moreover the graph of L is closed and hence also weakly closed. Consequently, the sequence (Lx_k) converges weakly to Lx. The assumption that $\mathcal{S}$ satisfies the Radon–Riesz property implies therefore that $Lx_k \to Lx$.

The assumption that $\operatorname{Ran} L$ is closed implies that the restriction of L to W (where $X = \operatorname{Ker} L \oplus W$) has a continuous inverse $L_W^{-1} \colon \operatorname{Ran} L \to W \subset X$. Denote now by P the projection onto $\operatorname{Ker} L$ along W and $P' = \operatorname{Id} - P$. Then in particular $L_W^{-1} L = P'$. Thus the convergence $Lx_k \to Lx$ and the continuity of L_W^{-1} on $\operatorname{Ran} L$ imply that $P'x_k \to P'x$.

Now the continuity of A implies that also $AP'x_k \to AP'x$. Together with the convergence $Ax_k \to Ax$, we therefore have that $APx_k \to APx$. Because the restriction of A to $\operatorname{Ker} L$ is boundedly invertible, this in turn implies that $Px_k \to Px$. Therefore, $x_k = Px_k + (\operatorname{Id} - P)x_k \to Px + (\operatorname{Id} - P)x = x$, which proves the assertion. □

4 Rates for Sparsity Regularisation

We now apply the results of the previous sections in order to derive convergence rates results for sparsity regularisation. We assume to that end that the target space Z of L has a structure similar to that of an ℓ^2-space and that the positively homogeneous functional $\mathcal{R}$ acts separately on the different components of Z. More precisely, we assume that $(V_\lambda)_{\lambda\in\Lambda}$ is a countable family of finite dimensional normed vector spaces and

$$Z = \left\{ z = (z_\lambda)_{\lambda\in\Lambda} : z_\lambda \in V_\lambda,\ \sum_{\lambda\in\Lambda} |z_\lambda|^2 < +\infty \right\} . \tag{11}$$

Moreover, we assume that we are given a family of convex, positively homogeneous functionals $\phi_\lambda \colon V_\lambda \to [0, +\infty]$ and that

$$\mathcal{S}(z) = \sum_{\lambda \in \Lambda} \phi_\lambda(z_\lambda) \, . \tag{12}$$

Note that we do not require that the spaces V_λ are isomorphic, but also in case they are, the norms on V_λ may be different for each index λ.

Assumption 4.1. Assume that the following conditions hold:

1. The space X is a reflexive Banach space, Y is a uniformly convex Banach space, and Z as given in (11).
2. The operator $A \colon X \to Y$ is bounded linear.
3. The regularisation term $\mathcal{R}$ has the form given by (10) and (12).
4. The mapping $L \colon D \subset X \to Z$ has a closed graph, closed range, and $\operatorname{Ker} L$ is finite dimensional.
5. We have $\operatorname{Ker} A \cap \operatorname{Ker} L = \{0\}$.
6. The functions ϕ_λ are lower semi-continuous, convex, and positively homogeneous. In addition, the domain of each function ϕ_λ is symmetric, that is, if $\phi_\lambda(z_\lambda) < +\infty$, then also $\phi_\lambda(-z_\lambda) < +\infty$.
7. There exists $d > 0$ such that $\phi_\lambda(v) \geq d|v|$ for every $\lambda \in \Lambda$ and $v \in V_\lambda$. ■

Lemma 4.2. *Let Assumption 4.1 hold. Then Tikhonov regularisation with $\mathcal{R}$ is well-posed, stable, and convergent.*

Proof. Similarly as in [19], one can show that $\mathcal{S}$ is weakly lower semi-continuous and coercive, and satisfies the Radon–Riesz property. Thus Remarks 3.3–3.5 show that Assumption 3.1 holds. Moreover, the assumptions of Lemma 3.7 hold, and thus the functional $\mathcal{R}$ satisfies the Radon–Riesz property. The assertion is therefore a consequence of Propositions 3.2 and 3.6. □

Remark 4.3. The well-posedness, stability, and convergence do not rely on the convexity and positive homogeneity of the regularisation term $\mathcal{R}$. They remain to hold, if we replace the functional $\mathcal{S}$ of (12) by a functional of the form

$$\mathcal{S}(z) = \sum_{\lambda \in \Lambda} \psi_\lambda(z_\lambda) \, ,$$

where the functions $\psi_\lambda \colon V_\lambda \to [0, +\infty]$ are lower semi-continuous satisfying $\psi(0) = 0$, $\lim_{|t| \to \infty} \psi(t) = +\infty$, and $\psi(t) \geq C|t|^2/(1 + |t|^2)$ for all $t \in V_\lambda$, where the constant $C > 0$ is independent of λ. Then, similarly as in [19], one can show that $\mathcal{S}$ is weakly lower semi-continuous and coercive, and satisfies the Radon–Riesz property. Thus, Propositions 3.2 and 3.6 are again applicable. Note that here we do not require that the functions ϕ_λ are convex. ■

For the derivation of linear convergence rates, it is necessary to describe the sub-differential of the regularisation functional $\mathcal{R}$. We have the inclusion

$$\partial\mathcal{R}(x^\dagger) \supset L^*\partial\mathcal{S}(Lx^\dagger) . \tag{13}$$

Moreover,

$$\begin{aligned}&\partial\mathcal{S}(Lx^\dagger)\\ &\quad= \Big\{\omega = (\omega_\lambda)_{\lambda\in\Lambda} : \omega_\lambda \in \partial\phi_\lambda((Lx^\dagger)_\lambda) \text{ for all } \lambda\in\Lambda \text{ and } \sum_{\lambda\in\Lambda}|\omega_\lambda|_*^2 < +\infty\Big\} .\end{aligned}$$

Here $|\omega_\lambda|_*$ denotes the norm on the dual V_λ^* of V_λ. Note that we need not have equality in (13), as we have not assumed that the mapping L is bounded and $\mathcal{S}$ continuous.

Now define for every $\lambda \in \Lambda$

$$\overline{\partial}\phi_\lambda(0) := \big\{w \in \partial\phi_\lambda(0) : (1+\varepsilon)w \notin \partial\phi_\lambda(0) \text{ for every } \varepsilon > 0\big\} ,$$

which, because ϕ is lower semi-continuous, positively homogeneous and coercive, is precisely the relative boundary of $\partial\phi(0)$. Because of the estimate $\phi_\lambda(v) \geq d|v|$ for every $\lambda \in \Lambda$ and $v \in V_\lambda$, this implies that $|w|_* \geq 1/d$ whenever $w \in \overline{\partial}\phi_\lambda(0)$. Note moreover that $\overline{\partial}\phi(0)$ can equivalently be written as

$$\overline{\partial}\phi_\lambda(0) = \bigcup_{v\in V_\lambda\setminus\{0\}} \partial\phi_\lambda(v) . \tag{14}$$

Now define for $\omega \in Z^*$ the index set

$$\Lambda_\omega := \big\{\lambda \in \Lambda : \omega_\lambda \notin \overline{\partial}\phi_\lambda(0)\big\}$$

and the space

$$Z_\omega := \big\{z = (z_\lambda)_{\lambda\in\Lambda} \in Z : z_\lambda = 0 \text{ whenever } \lambda \in \Lambda_\omega\big\} .$$

Because $|\omega_\lambda|_* \geq 1/d$ whenever $\omega_\lambda \in \overline{\partial}\phi_\lambda(0)$ and $\omega \in Z^*$, implying that $\sum_\lambda |\omega_\lambda|_*^2 < +\infty$, it follows that $\Lambda \setminus \Lambda_\omega$ is a finite set and, consequently, Z_ω a finite dimensional subspace of Z.

Theorem 4.4. *Let Assumption 4.1 hold. Assume that there exists a source element $\omega = (\omega_\lambda)_{\lambda\in\Lambda} \in Z^*$ such that $\omega_\lambda \in \partial\phi_\lambda\big((Lx^\dagger)_\lambda\big)$ for every $\lambda \in \Lambda$ and* $\operatorname{Ran} A^* \cap L^*\omega \neq \emptyset$. *Assume moreover that the restriction of A to the preimage $L^{-1}Z_\omega$ of Z_ω is injective. Then we have for a parameter choice $\alpha \sim \delta$ that*

$$\mathcal{R}(x_\alpha^\delta - x^\dagger) = O(\delta) \qquad \text{as } \delta \to 0 .$$

Proof. This result follows by applying Theorem 2.6 with $\tilde{K} = (L^{-1}Z_\omega) \cap \operatorname{Dom}\mathcal{R}$. To that end note first that the definition of ω implies that $L^*\omega \in \partial\mathcal{R}(x^\dagger)$. Moerover, the finite dimension of Z_ω and the assumption that $\dim(\operatorname{Ker} L) < +\infty$ imply that $L^{-1}Z_\omega$ is a finite dimensional subspace of X. In view of Remark 2.7, this shows (6).

Next we verify that $(L^{-1}Z_\omega) \cap \operatorname{Dom}\mathcal{R}$ contains the cone $K_{L^*\omega}$ generated by the sub-gradient element $L^*\omega \in \partial\mathcal{R}(x^\dagger)$. That is, we show that

$$K_{L^*\omega} := \big\{x \in X : L^*\omega \in \partial\mathcal{R}(x)\big\} \subset (L^{-1}Z_\omega) \cap \operatorname{Dom}\mathcal{R} . \tag{15}$$

Note first that by definition $K_{L^*\omega} \subset \operatorname{Dom} \mathcal{R} \subset D$. Moreover, the symmetry of the domains of the functions ϕ_λ implies that also $-x^\dagger \in \operatorname{Dom} \mathcal{R}$, and therefore $\operatorname{Dom} \mathcal{R} = x^\dagger + \operatorname{Dom} \mathcal{R}$. Therefore (15) is equivalent to the inclusion $L(K_{L^*\omega}) \subset Z_\omega$. Now recall that $x \in K_{L^*\omega}$, if and only if $\langle L^*\omega, x\rangle = \mathcal{R}(x)$. This, however, is equivalent to the equation $\langle \omega, Lx\rangle = \mathcal{S}(Lx)$, which in turn implies that $\omega \in \partial\mathcal{S}(Lx)$. Let now $\lambda \in \Lambda$ be such that $\omega_\lambda \notin \overline{\partial}\phi_\lambda(0)$. Then (14) and the fact that $\omega_\lambda \in \partial\phi_\lambda((Lx)_\lambda)$ imply that $(Lx)_\lambda = 0$, which proves (15).

In order to be able to apply Theorem 2.6, it remains to show (7). To that end, define for $\lambda \in \Lambda$

$$\varepsilon_\lambda := \sup\bigl\{\varepsilon \geq 0 : (\omega_\lambda + \varepsilon_\lambda \partial\phi_\lambda(0)) \subset \partial\phi_\lambda(0)\bigr\} .$$

Then we have $\lambda \in \Lambda_\omega$ if and only if $\varepsilon_\lambda = 0$. Now assume that $\lambda \notin \Lambda_\omega$ is such that $|\omega_\lambda|_* \leq 1/(2d)$. Then we also have that $2\omega_\lambda \in \partial\phi_\lambda(0)$. Consequently, the convexity of $\partial\phi_\lambda(0)$ implies that $\omega_\lambda + \frac{1}{2}\partial\phi_\lambda(0) \subset \partial\phi_\lambda(0)$, showing that, in this case, $\varepsilon_\lambda \geq 1/2$. Because $|\omega_\lambda|_* > 1/(2d)$ for at most finitely many $\lambda \in \Lambda$, this implies that the set of indices λ satisfying $0 < \varepsilon_\lambda < 1/2$ is finite. Therefore

$$\tilde{c}_0 := \inf\bigl\{\varepsilon_\lambda : \lambda \notin \Lambda_\omega\bigr\} > 0 .$$

Noting that

$$\mathcal{N}(Z_\omega) = \bigl\{z = (z_\lambda)_{\lambda\in\Lambda} \in Z : z_\lambda = 0 \text{ whenever } \lambda \notin \Lambda_\omega\bigr\} ,$$

we see that

$$\tilde{c}_0 = \max\bigl\{c \geq 0 : (\omega + c\partial\mathcal{S}(0) \cap \mathcal{N}(Z_\omega)) \subset \partial\mathcal{S}(0)\bigr\} . \tag{16}$$

Let $\xi \in \mathcal{N}(L^{-1}Z_\omega) \cap \operatorname{Ran} L^*$ and let $\eta \in Z^*$ be such that $L^*\eta = \xi$. Then $0 = \langle \xi, \tilde{x}\rangle = \langle L^*\eta, \tilde{x}\rangle = \langle \eta, L\tilde{x}\rangle$ for every $\tilde{x} \in L^{-1}Z_\omega$, showing that $\eta \in \mathcal{N}(Z_\omega \cap \operatorname{Ran} L)$. Now note that

$$\mathcal{N}(Z_\omega \cap \operatorname{Ran} L) = \mathcal{N}(Z_\omega) \oplus (Z_\omega^* \cap \mathcal{N}(\operatorname{Ran} L)) = \mathcal{N}(Z_\omega) \oplus (Z_\omega^* \cap \operatorname{Ker} L^*) .$$

This shows that there exists $\tilde{\eta} \in \mathcal{N}(Z_\omega)$ such that $L^*\tilde{\eta} = \xi$ and consequently $\mathcal{N}(L^{-1}Z_\omega) \cap \operatorname{Ran} L^* \subset L^*(\mathcal{N}(Z_\omega) \cap D')$, where $D' := \operatorname{Dom} L^*$. We therefore obtain from (16) that

$$L^*\omega + \tilde{c}_0 L^*(\partial\mathcal{S}(0) \cap D') \cap \mathcal{N}(L^{-1}Z_\omega) \subset L^*(\partial\mathcal{S}(0) \cap D') \subset \partial\mathcal{R}(0) . \tag{17}$$

Next we show that $C := L^*(\partial\mathcal{S}(0) \cap D') \cap \mathcal{N}(L^{-1}Z_\omega)$ is a dense subset of $\partial\mathcal{R}(0) \cap \mathcal{N}(L^{-1}(Z_\omega))$. Assume to the contrary that $\xi \in \partial\mathcal{R}(0) \cap \mathcal{N}(L^{-1}(Z_\omega)) \setminus \overline{C}$. Then there exists $x \in X$ such that

$$\begin{aligned} \mathcal{R}(x) \geq \langle \xi, x\rangle > \sup\bigl\{\langle \zeta, x\rangle : \zeta \in C\bigr\} \\ = \sup\bigl\{\langle L^*\eta, x\rangle : \eta \in \partial\mathcal{S}(0) \cap D' \cap \mathcal{N}(Z_\omega)\bigr\} . \end{aligned} \tag{18}$$

Note moreover that we may assume without loss of generality that $x \in L^{-1}(Z_\omega)$, as $\xi \in \mathcal{N}(L^{-1}(Z_\omega))$. Thus $\langle \eta, Lx\rangle = 0$ for every $\eta \in \mathcal{N}(Z_\omega)$, implying that

$$\begin{aligned} &\sup\bigl\{\langle L^*\eta, x\rangle : \eta \in \partial\mathcal{S}(0) \cap D' \cap \mathcal{N}(Z_\omega)\bigr\} \\ &\qquad = \sup\bigl\{\langle \eta, Lx\rangle : \eta \in \partial\mathcal{S}(0) \cap D' \cap \mathcal{N}(Z_\omega)\bigr\} \\ &\qquad = \sup\bigl\{\langle \eta, Lx\rangle : \eta \in \partial\mathcal{S}(0) \cap D'\bigr\} \\ &\qquad = \mathcal{S}(Lx) . \end{aligned}$$

Because $\mathcal{S}(Lx) = \mathcal{R}(x)$, this contradicts (18), and therefore $C \subset \partial\mathcal{R}(0) \cap \mathcal{N}(L^{-1}(Z_\omega))$ is dense. Consequently (17) and the fact that $\partial\mathcal{R}(0)$ is closed imply that also

$$L^*\omega + \tilde{c}_0 \partial\mathcal{R}(0) \cap \mathcal{N}(L^{-1}(Z_\omega)) \subset \partial\mathcal{R}(0) .$$

Since $\tilde{c}_0 > 0$, this shows that the assumptions of Theorem 2.6 are satisfied, which therefore proves the assertion. □

5 Joint Sparsity

In this section, we apply Theorem 4.4 for deriving convergence rates for sparse regularisation of vector valued data. In the standard, scalar valued case, sparse regularisation tries to reconstruct sequences with a small (finite) number of non-zero coefficients. In the setting of joint sparsity, the goal is in principle the same, but the coefficients of the sequences are themselves vectors and one does not require sparsity of every single coefficient vector. In contrast, one typically wants to obtain sparse sequences of dense vectors. A standard method for obtaining such results is the usage of a regularisation term that consists of the sum of the Euclidean norms of the coefficient vectors (see for instance [3, 18, 29, 32]). Below, we also consider the setting where different norms than only the Euclidean one are used for the penalisation of the coefficient vectors.

We assume that the space X is a vector valued ℓ^2-space, that is,

$$X = \Big\{ x = (x_\lambda)_{\lambda\in\Lambda} : x_\lambda \in \mathbb{R}^n,\ \sum_{\lambda\in\Lambda} |x_\lambda|^2 < +\infty \Big\}$$

for some $n \in \mathbb{N}$ and a countable index set Λ. Here $|x_\lambda|$ denotes the Euclidean norm of the vector $x_\lambda \in \mathbb{R}^n$. Moreover, we assume that the regarisation term $\mathcal{R}$ is defined by some convex and positively homogeneous function $\phi\colon \mathbb{R}^n \to [0, +\infty]$ with symmetric domain satisfying $\phi(v) = 0$ if and only if $v = 0$ via

$$\mathcal{R}(x) = \sum_\lambda \phi(x_\lambda) \qquad \text{for } x = (x_\lambda)_{\lambda\in\Lambda} \text{ with } x_\lambda \in \mathbb{R}^n .$$

Then

$$\partial\mathcal{R}(x^\dagger) = \Big\{ \xi = (\xi_\lambda)_{\lambda\in\Lambda} : \xi_\lambda \in \partial\phi(x_\lambda^\dagger) \text{ for all } \lambda \in \Lambda \text{ and } \sum_{\lambda\in\Lambda} |\xi_\lambda|^2 < +\infty \Big\} .$$

Now define as in Section 4

$$\overline{\partial}\phi(0) := \big\{ \xi \in \partial\phi(0) : (1+\varepsilon)\omega \notin \partial\phi(0) \text{ for every } \varepsilon > 0 \big\} .$$

and

$$\Lambda_\xi := \big\{ \lambda \in \Lambda : \xi_\lambda \in \overline{\partial}\phi(0) \big\}$$

for $\xi \in X$.

Proposition 5.1. *Assume that there exists $\xi \in \partial\mathcal{R}(x^\dagger) \cap \operatorname{Ran} A^*$ such that the restriction of A to*

$$\ell^2(\Lambda_\xi; \mathbb{R}^n) := \big\{ x = (x_\lambda)_{\lambda\in\Lambda} \in X : x_\lambda = 0 \textit{ for } \lambda \notin \Lambda_\omega \big\}$$

is injective. Then we have for a parameter choice $\alpha \sim \delta$ that

$$\mathcal{R}(x_\alpha^\delta - x^\dagger) = O(\delta) \qquad \textit{as } \delta \to 0 .$$

In particular, $\|x_\alpha^\delta - x^\dagger\| = O(\delta)$.

Proof. This is a special case of Theorem 4.4 with $Z = X$ and $L = \mathrm{Id}\colon X \to Z$. Note that the coercivity condition $\phi(v) \geq d|v|$ for some $d > 0$ follows from the assumption that 0 is the unique minimiser of ϕ. □

Remark 5.2. The convergence result of Proposition 5.1 can easily be extended to regularisation terms $\mathcal{R}$ of the form

$$\mathcal{R}(x) = \sum_{\lambda \in \Lambda} \phi_\lambda(x_\lambda) ,$$

where the convex and positively homogenous functions ϕ_λ with symmetric domains in addition depend on the index λ. There one only has the additional requirement that the family of functions $(\phi_\lambda)_{\lambda\in\Lambda}$ has to be equicoercive, that is, there exists some $d > 0$ such that $\phi_\lambda(v) \geq d|v|$ for all $v \in \mathbb{R}^n$ and $\lambda \in \Lambda$.

A similar extension is possible for the setting where the dimension of the coefficients of $x = (x_\lambda)_{\lambda\in\Lambda}$ depends on the index λ. That is,

$$X = \bigl\{ x = (x_\lambda)_{\lambda\in\Lambda} : x_\lambda \in V_\lambda,\ \sum_{\lambda\in\Lambda} |x_\lambda|_{V_\lambda}^2 < \infty \bigr\} ,$$

where $(V_\lambda)_{\lambda\in\Lambda}$ is a family of finite dimensional vector spaces. Note that we do not need any uniform bound on the dimension of the spaces; it is possible that $\sup_\lambda \dim(V_\lambda) = +\infty$, as long as the dimension of each space V_λ is finite. ■

6 Group Sparsity

We now return to the case of scalar valued data, but assume that the underlying index set has some known structure that allows us to assign the indices to different groups, and we have the a–priori knowledge that only a small number of groups contribute to the data. Then, similarly as for joint sparsity, it makes sense to penalise the Euclidean norms within each group and then sum over all the different groups (see for instance [15, 31]). Formally, we can model this situation by assuming that $X = \ell^2(\Sigma)$ for some countable index set Σ and that we are given another countable index set Λ and a mapping $\pi\colon \Lambda \to 2^\Sigma$ assigning to each $\lambda \in \Lambda$ a group of indices in Σ. Then we would choose the regularisation term

$$\mathcal{R}(x) = \sum_{\lambda\in\Lambda} \Bigl(\sum_{\sigma\in\pi(\lambda)} x_\sigma^2 \Bigr)^{1/2} .$$

In case the groups are disjoint, that is, $\pi(\lambda) \cap \pi(\lambda') = \emptyset$ whenever $\lambda \neq \lambda'$, this setting is equivalent to that of joint sparsity considered in the second part of Remark 5.2. There are, however, distinct differences when the groups are allowed to overlap. Such overlapping groups have been recently used in [25, 26].

We say that the group assignment π is locally finite, if

- for every $\lambda \in \Lambda$ the set $\pi(\lambda)$ is finite, and

- for every $\sigma \in \Sigma$ the set $\pi^{-1}(\sigma) = \{\lambda \in \Lambda : \sigma \in \pi(\lambda)\}$ is finite.

Proposition 6.1. *Assume that π is locally finite and $\pi^{-1}(\sigma) \neq \emptyset$ for every $\sigma \in \Sigma$. Assume moreover that there exist $\omega_{\lambda,\sigma} \in \mathbb{R}$, $\lambda \in \Lambda$, $\sigma \in \pi(\lambda)$, such that*

$$\omega_{\lambda,\sigma} = x^\dagger_\sigma \Big(\sum_{\tau\in\pi(\lambda)} {x^\dagger_\tau}^2 \Big)^{-1/2}$$

whenever $\sigma \in \pi(\lambda)$ and there exists $\tau \in \pi(\lambda)$ such that $x^\dagger_\tau \neq 0$,

$$\sum_{\tau\in\pi(\lambda)} \omega^2_{\lambda,\tau} \leq 1$$

whenever $\lambda \in \Lambda$ is such that $x^\dagger_\sigma = 0$ for every $\sigma \in \pi^{-1}(\lambda)$, and

$$\sum_{\lambda\in\Lambda} \sum_{\sigma\in\pi(\lambda)} \omega^2_{\lambda,\sigma} < +\infty\ .$$

Define $\xi \in \ell^2(\Sigma)$ by

$$\xi_\sigma := \sum_{\lambda\in\pi^{-1}(\sigma)} \omega_{\lambda,\sigma}\ ,$$

and denote

$$\Sigma_\omega := \Big\{ \sigma \in \Sigma : \sigma \in \pi(\lambda) \text{ for some } \lambda \in \Lambda \text{ such that } \sum_{\tau\in\pi(\lambda)} \omega^2_{\lambda,\tau} = 1 \Big\}\ .$$

If $\xi \in \operatorname{Ran}(A^*)$ and the restriction of A to

$$\ell^2(\Sigma_\omega) := \big\{ x = (x_\sigma)_{\sigma\in\Sigma} \in \ell^2(\Sigma) : x_\sigma = 0 \text{ for all } \sigma \notin \Sigma_\omega \big\}$$

is injective, then

$$\mathcal{R}(x^\delta_\alpha - x^\dagger) = O(\delta) \qquad \text{as } \delta \to 0\ .$$

In particular, $\|x^\delta_\alpha - x^\dagger\| = O(\delta)$.

Proof. Define

$$Z := \Big\{ z = (z_\lambda)_{\lambda\in\Lambda} : z_\lambda \in \mathbb{R}^{|\pi(\lambda)|},\ \sum_{\lambda\in\Lambda} |z_\lambda|^2 < \infty \Big\}\ ,$$

let

$$D := \Big\{ x = (x_\sigma)_{\sigma\in\Sigma} : \sum_{\lambda\in\Lambda} \sum_{\sigma\in\pi(\lambda)} x^2_\sigma < +\infty \Big\}\ ,$$

and define the possibly unbounded mapping $L\colon D \subset X \to Z$ by $(Lx)_\lambda = (x_\sigma)_{\sigma\in\pi(\lambda)}$. Because $\pi^{-1}(\sigma) \neq \emptyset$ for every $\sigma \in \Sigma$, the mapping L is injective. Moreover, it is easy to see that L has a closed graph and a closed range.

The assumptions on ω imply that $\omega \in Z^*$ and $L^*\omega \in \partial\mathcal{R}(x^\dagger)$. Finally, the injectivity of A on $\ell^2(\Sigma_\omega)$ directly translates to the injectivity condition required in Theorem 4.4. Thus said theorem is applicable to this situation, and the linear convergence rate follows. □

Remark 6.2. Though, strictly speaking, the local finiteness of the assignment π is not required for the convergence rates result, it makes sense to postulate it nevertheless. One the one hand, it ensures that every finite sequence $x \in \ell^2(\Sigma)$ is contained in the domain of $\mathcal{R}$. On the other hand, it guarantees that the sub-differential of $\mathcal{R}$ at each finite sequence is non-empty. In fact, these two conditions together are equivalent to the local finiteness of π. ■

7 Discrete Total Variation Regularisation

The assumption of the finite dimension of the space in which the true solution is supported indicates that it is not possible to apply Theorems 4.4 or 2.6 to regularisation methods based on (possibly higher order) total variation. It does not exclude the application to *discretisations* of total variation regularisation, though.

Exemplarily, we will discuss the setting of first order discrete total variation regularisation in the two-dimensional case. Here we assume that we are given some uniform rectangular grid $I \subset \mathbb{Z}^2$ and define a discrete gradient $\nabla \colon \ell^2(I) \to \ell^2(I;\mathbb{R}^2)$ by (cf. the discretisation of the total variation proposed in [8])

$$(\nabla u)_{i,j} := \begin{cases} (u_{i+1,j} - u_{i,j}; u_{i,j+1} - u_{i,j})\,, & \text{if } (i+1,j) \in I \text{ and } (i,j+1) \in I\,,\\ (u_{i+1,j} - u_{i,j}; 0)\,, & \text{if } (i+1,j) \in I \text{ and } (i,j+1) \notin I\,,\\ (0; u_{i,j+1} - u_{i,j})\,, & \text{if } (i+1,j) \notin I \text{ and } (i,j+1) \in I\,,\\ (0;0)\,, & \text{if } (i+1,j) \notin I \text{ and } (i,j+1) \notin I\,. \end{cases}$$

Moreover, we define

$$\mathcal{R}(u) := \|\nabla u\| := \sum_{(i,j)\in I} |(\nabla u)_{i,j}|$$

with $|v| := \sqrt{v_1^2 + v_2^2}$ for $v = (v_1, v_2) \in \mathbb{R}^2$.

We assume in addition that the index set I is connected in the sense that whenever (i,j), $(\tilde{\imath},\tilde{\jmath}) \in I$, there exists a path $\big((i_0,j_0),(i_1,j_1),\ldots,(i_n,j_n)\big) \subset I$ with $(i_0,j_0) = (i,j)$, $(i_n,j_n) = (\tilde{\imath},\tilde{\jmath})$, and $|i_{k+1} - i_k| + |j_{k+1} - j_k| = 1$ for every k. This assumption implies that $\operatorname{Ker}\nabla$ consists either of all constant sequences in case $|I| < +\infty$, or $\operatorname{Ker}\nabla = \{0\}$ if I is an infinite set. In particular, $\operatorname{Ker}\nabla$ is finite dimensional.

Finally, note that the adjoint of the discrete gradient is the discrete divergence defined as (cf. [8]) $\nabla^* = \operatorname{div} \colon \ell^2(I;\mathbb{R}^2) \to \ell^2(I)$,

$$(\operatorname{div} V)_{i,j} = \begin{cases} V^{(1)}_{i,j} - V^{(1)}_{i-1,j} & \text{if } (i\pm 1,j) \in I\\ V^{(1)}_{i,j} & \text{if } (i-1,j) \notin I\\ -V^{(1)}_{i-1,j} & \text{if } (i+1,j) \notin I\\ 0 & \text{if } (i\pm 1,j) \notin I \end{cases} + \begin{cases} V^{(2)}_{i,j} - V^{(2)}_{i,j-1} & \text{if } (i,j\pm 1) \in I\\ V^{(2)}_{i,j} & \text{if } (i,j-1) \notin I\\ -V^{(2)}_{i,j-1} & \text{if } (i,j+1) \notin I\\ 0 & \text{if } (i,j\pm 1) \notin I \end{cases}.$$

Proposition 7.1. *Assume that there exists $V = (V_{i,j})_{(i,j)\in I} \in \ell^2(I;\mathbb{R}^2)$ such that $V_{i,j} = (\nabla u^\dagger)_{i,j}/|(\nabla u^\dagger)_{i,j}|$ whenever $(\nabla u^\dagger)_{i,j} \neq 0$ and $|V_{i,j}| \leq 1$ else, and $\operatorname{div} V \in \operatorname{Ran} A^*$. Denote $J := \big\{(i,j) \in I : |V_{i,j}| < 1\big\}$. If the restriction of A to*

$$S := \big\{u \in \ell^2(I) : (\nabla u)_{i,j} = 0 \text{ for every } (i,j) \in J\big\}$$

is injective, then we have for a parameter choice $\alpha \sim \delta$ that

$$\mathcal{R}(u_\alpha^\delta - u^\dagger) = \|\nabla(u_\alpha^\delta - u^\dagger)\| = O(\delta) \qquad \text{as } \delta \to 0\,.$$

Proof. This is a direct consequence of Theorem 4.4. □

8 Conclusion

This paper introduces mathematical tools that can be used for extending existing results on linear convergence rates for Tikhonov regularisation with sparsity promoting regularisation terms to more general settings. Exemplarily, these tools are applied in three different settings—joint sparsity, group sparsity, and discrete total variation regularisation—where, up to now, no comparable convergence rates results have 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] R. Acar and C. R. Vogel. Analysis of bounded variation penalty methods for ill-posed problems. *Inverse Probl.*, 10(6):1217–1229, 1994.

[2] G. Aubert and L. Vese. A variational method in image recovery. *SIAM J. Numer. Anal.*, 34(5):1948–1979, 1997.

[3] E. van den Berg and M.P. Friedlander. Theoretical and empirical results for recovery from multiple measurements. *IEEE Trans. Inf. Theory*, 56(5):2516–2527, 2010.

[4] K. Bredies. A forward-backward splitting algorithm for the minimization of non-smooth convex functionals in Banach space. *Inverse Probl.*, 25(1):015005, 20, 2009.

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

[6] 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.

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

[8] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vision*, 20(1–2):89–97, 2004.

[9] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numer. Math.*, 76(2):167–188, 1997.

[10] 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.

[11] G. Dal Maso. *An Introduction to* Γ*-Convergence*, volume 8 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser, 1993.

[12] 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.

[13] 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.

[14] D. L. Donoho, M. Elad, and V. N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inf. Theory*, 52(1):6–18, 2006.

[15] Y. C. Eldar and M. Mishali. Robust recovery of signals from a structured union of subspaces. *IEEE Trans. Inf. Theory*, 55(11):5302–5316, 2009.

[16] 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.

[17] M. Figueiredo, R. Nowak, and S. Wright. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. *IEEE J. Sel. Topics Signal Process.*, 1(4):586–598, 2007.

[18] M. Fornasier and H. Rauhut. Recovery algorithms for vector-valued data with joint sparsity constraints. *SIAM J. Numer. Anal.*, 46(2):577–613, 2008.

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

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

[21] 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, 2009.

[22] M. Grasmair, M. Haltmeier, and O. Scherzer. Sparsity in inverse geophysical problems. In W. Freeden, M. Z. Nashed, and T. Sonar, editors, *Handbook of Geomathematics*, pages 763–784. Springer Berlin Heidelberg, 2010.

[23] M. Grasmair, M. Haltmeier, and O. Scherzer. Necessary and sufficient conditions for linear convergence of ℓ^1-regularization. *Comm. Pure Appl. Math.*, 64(2):161–182, 2011.

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

[25] L. Jacob, G. Obozinski, and J.-Ph. Vert. Group lasso with overlaps and graph lasso. In L. Bottou and M. Littman, editors, *Proceedings of the 26th International Conference on Machine Learning*, pages 433–440, Montreal, June 2009. Omnipress.

[26] J. Liu and J. Ye. Fast overlapping group lasso. Preprint, Arizona State University, Tempe, AZ, 2010. arXiv:1009.0306.

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

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

[29] F. Pitolli and G. Bretti. An iterative algorithm with joint sparsity constraints for magnetic tomography. In M. Dæhlen, M. Floater, T. Lyche, J.-L. Merrien, K. Mørken, and L. Schumaker, editors, *Mathematical Methods for Curves and Surfaces*, volume 5862 of *Lecture Notes in Computer Science*, pages 316–328. Springer Berlin / Heidelberg, 2010.

[30] 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.

[31] M. Stojnic, F. Parvaresh, and B. Hassibi. On the reconstruction of block-sparse signals with an optimal number of measurements. *IEEE Trans. Inf. Theory*, 57(8):3075–3085, 2009.

[32] G. Teschke and R. Ramlau. An iterative algorithm for nonlinear inverse problems with joint sparsity constraints in vector-valued regimes and an application to color image inpainting. *Inverse Probl.*, 23(5):1851–1870, 2007.

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