Iterative Thresholding Algorithms

Massimo Fornasier* and Holger Rauhut†

March 8, 2007

Abstract

This article provides a variational formulation for hard and firm thresholding. A related functional can be used to regularize inverse problems by sparsity constraints. We show that a damped hard or firm thresholded Landweber iteration converges to its minimizer. This provides an alternative to an algorithm recently studied by the authors. We prove stability of minimizers with respect to the parameters of the functional and its regularization properties by means of Γ-convergence. All investigations are done in the general setting of vector-valued (multi-channel) data.

Key words: linear inverse problems, joint sparsity, thresholded Landweber iterations, variational calculus on sequence spaces, Γ-convergence

AMS subject classifications: 65J22, 65K10, 65T60, 90C25, 52A41, 49M30, 47J30

1 Introduction

Thresholding is a simple technique for denoising signals and images. When the signal is represented in terms of a suitable basis (for instance a wavelet basis) small coefficients are set to zero and larger coefficients above some threshold are possibly shrinked. Therefore, thresholding (or shrinkage) usually produces signals that are sparse, i.e., that have only a small number of non-zero coefficient. So it works particular well if the original noise free signal can be well-approximated by a sparse one.

Since the seminal papers [15, 18, 19] *soft* and *hard* thresholding operators have been extensively studied. While both have been used indifferently in the practice, from a theoretical point of view the first attracted most of the attention. In fact [7] established a variational formulation for denoising by ℓ_1 penalization, which results in simple soft-thresholding. This interpretation has caught much attention due to its similarity and near-equivalence to the well-known Rudin-Osher-Fatemi de-noising model [29] based on total variation minimization. While soft-thresholding is a very simple operation, total variation minimization

*Princeton University, Program in Applied and Computational Mathematics, Fine Hall, Washington Road, Princeton NJ 08544-1000, U.S.A., E-mail: mfornasi@math.princeton.edu.
M. F. acknowledges the financial support provided by the European Union's Human Potential Programme under contract MOIF-CT-2006-039438.
†Numerical Harmonic Analysis Group, Department of Mathematics, University of Vienna, Nordbergstrasse 15, A-1090 Vienna. E-mail: holger.rauhut@univie.ac.at.
H. R. was supported by the Individual Marie-Curie Fellowship MEIF-CT 2006-022811.

requires the solution of a degenerate PDE. Although the latter problem has stimulated interesting results both in theory and practice, see for instance [8], soft-thresholding still remains a very efficient and simple numerical strategy. The interested reader is referred to [13] for a recent and interesting comparison of these two methods.

As one of the main contributions of this paper we establish a connection between hard-thresholding and the minimizer of a *convex* functional. To our knowledge this is the first convex variational interpretation of hard-thresholding. We had actually introduced the corresponding functional already in [24], but while later investigating the role of its parameters we realized the surprising fact that its minimizer is connected to hard-thresholding in a special case. Varying parameters allows further to "interpolate" between soft and hard thresholding. It turns out that the resulting generalized thresholding operators coincide with the ones introduced in [25] under the name of *firm shrinkage*.

In many situations the signal is not given explicitly but implicitly by an operator equation or linear inverse problem with possibly noisy data. We refer for instance to deconvolution and super-resolution problems [14, 16, 31], image recovery and enhancing [12, 21], problems arising in geophysics and biomedical imaging [23, 26], statistical estimation [20, 34], or to compressed sensing [3, 5, 17, 28]. A strategy which has recently become popular is to regularize these reconstruction problems by ℓ_1-constraints. Unfortunately, the minimizer of the resulting functional can no longer be computed explicitly. Several authors have proposed an iterative soft thresholding algorithm to approximate the solution [22, 30, 31, 20], and in [11] its convergence was proved. On the basis of these recent achievements for the solution of inverse problems with sparsity constraints, several articles appeared with further generalizations [2, 27, 32, 33].

In [24] we proposed a new functional involving a weighted ℓ_1-norm with adaptive weights, i.e., the weights are variables of the functional as well. The minimizer of this functional is suggested as the regularized solution of the corresponding linear inverse problem. This functional again promotes sparsity with respect to a suitable basis or frame, and as we already mentioned above it is related to a (damped) hard or firm thresholding operator. However, we note that unless the operator in the inverse problem is unitary, the corresponding functional is only convex if we introduce a certain quadratic term which leads to an additional damping in the hard or firm thresholding operator.

Recently, similar approaches to adaptive weights appeared also in sparse statistical estimation, see e.g. [37] and references therein.

Our work in [24] was originally motivated by vector valued (multi-channel) problems, where each vector component possesses a sparse expansion with respect to the same frame, and additionally the different components obey the same sparsity pattern, i.e., have their non-zero components at the same locations. Color images are a typical example of such vector valued data with coupled components. Common sparsity patterns can be modelled with weighted ℓ_1-norms of componentwise ℓ_q norms of the coefficients where typically $q > 1$, see also [6, 35]. All results in this article will be derived in this general multi-channel setting, although for many applications the mono-channel case will be sufficient. In particular, we will also derive (several) generalizations of the hard and firm thresholding operator to the vector-valued case (while corresponding soft thresholding operators were already computed in [24]).

In [24] we suggested an algorithm for the minimization of our new functional. It consists in alternating a minimization with respect to the frame coefficients and a minimization with respect to the weights in the ℓ_1-norm. The former is done via a (possibly damped) soft thresholded Landweber iteration scheme, which in the monochannel case coincides with the algorithm analyzed by Daubechies et al. in [11]. The minimizing weights (for fixed coefficients) can be computed explicitly. The convergence of this two-step algorithm is shown in [24].

Realizing the connection of our functional to (damped) hard and firm thresholding, it is natural to ask whether the corresponding thresholded Landweber iteration converges as well to its minimizer. Under certain conditions on the parameters ensuring (strict) convexity of the functional, we prove such convergence. Compared with our first two-step algorithm the new algorithm clearly has the advantage of providing a single iteration scheme, and we expect that it will have faster convergence in practice (although this issue is postponed to later investigations). Our variational formulation generates a new family of iterative damped thresholding algorithms as a natural extension of soft and hard thresholding. We further note that [4, 36] investigate properties of iterative pure hard thresholding schemes, in particular, [4] shows its convergence. However, for non-trivial operators it seems that one can only associate a *non-convex* functional to the pure non-damped hard-thresholding operator, so the algorithm provides only a local minimum which may differ for different initial points.

We also discuss the dependence of the minimizers on the parameters and their stability properties. In particular, we show that the minimizers of our functional weakly converge to the minimizer of the ℓ_1-regularized functional analyzed in [11] for certain limits of the parameters. The proof requires the use of the Γ-convergence machinery [10]. As a further corollary we prove regularization results, which correspond again to certain limits of the parameters. The variational techniques employed in the last part of the paper reflect very much the similarities between our functional with the one proposed by Chambolle in [6] as the discrete counterpart of the Ambrosio-Tortorelli approximation [1] of the Mumford-Shah functional.

The paper is organized as follows. Section 2 introduces notation and our functional. Further, we recall the double-minimization algorithm in [24] and the corresponding convergence result. Section 3 discusses the connection to hard and firm thresholding operators, and derives their generalization to the vector valued case. Section 4 is devoted to the convergence proof of the thresholded Landweber iteration to minimizers of our original functional. The dependence of minimizer on the parameters will be discussed in Section 5.

2 Motivation

2.1 Some notation

Before starting our discussion let us briefly introduce some of the spaces we will use in the following. For some countable index set Λ we denote by $\ell_p = \ell_p(\Lambda)$, $1 \leq p \leq \infty$ the space of real sequences $u = (u_\lambda)_{\lambda \in \Lambda}$ with norm

$$\|u\|_p = \|u|\ell_p\| := \left(\sum_{\lambda \in \Lambda} |u_\lambda|^p \right)^{1/p}, \quad 1 \leq p < \infty$$

and $\|u\|_\infty := \sup_{\lambda\in\Lambda} |u_\lambda|$ as usual. If (v_λ) is a sequence of positive weights then we define the weighted spaces $\ell_{p,v} = \ell_{p,v}(\Lambda) = \{u, (u_\lambda v_\lambda) \in \ell_p(\Lambda)\}$ with norm

$$\|u\|_{p,v} \;=\; \|u|\ell_{p,v}\| \;=\; \|(u_\lambda v_\lambda)\|_p \;=\; \left(\sum_{\lambda\in\Lambda} v_\lambda^p |u_\lambda|^p)\right)^{1/p}$$

(with the usual modification for $p = \infty$). If the entries u_λ are actually vectors in a Banach space X with norm $\|\cdot\|_X$ then we denote

$$\ell_{p,v}(\Lambda, X) \;:=\; \{(u_\lambda)_{\lambda\in\Lambda}, u_\lambda \in X, (\|u_\lambda\|_X)_{\lambda\in\Lambda} \in \ell_{p,v}(\Lambda)\}$$

with norm $\|u|\ell_{p,v}(\Lambda, X)\| \;=\; \|(\|u_\lambda\|_X)_{\lambda\in\Lambda}|\ell_{p,v}(\Lambda)\|$. Usually X will be $\mathbb{R}^M$ endowed with the Euclidean norm, or the M-dimensional space ℓ_q^M, i.e., $\mathbb{R}^M$ endowed with the ℓ_q-norm. By $\mathbb{R}_+$ we denote the non-negative real numbers.

2.2 A functional modelling joint sparsity

Let $\mathcal{K}$ and $\mathcal{H}_j$, $j = 1, \dots, N$, be (separable) Hilbert spaces and $A_{\ell,j} : \mathcal{K} \to \mathcal{H}_j$, $j = 1, \dots, M$, $\ell = 1, \dots, N$ some bounded linear operators. Assume we are given data $g_j \in \mathcal{H}_j$,

$$g_j \;=\; \sum_{\ell=1}^{M} A_{\ell,j} f_\ell, \quad j = 1, \dots, N.$$

Our main task consists in reconstructing the (unknown) elements $f_\ell \in \mathcal{K}$, $\ell = 1, \dots, M$.

To address the formulation of an algorithm to recover the vector components $f_\ell \in \mathcal{K}$, $\ell = 1, \dots, M$, assume that we have given a suitable frame $(\psi_\lambda)_{\lambda\in\Lambda} \subset \mathcal{K}$ indexed by a countable set Λ. This means that there exist constants $C_1, C_2 > 0$ such that

$$C_1\|f\|_{\mathcal{K}}^2 \le \sum_{\lambda\in\Lambda} |\langle f, \psi_\lambda\rangle|^2 \le C_2\|f\|_{\mathcal{K}}^2 \qquad \text{for all } f \in \mathcal{K}. \tag{2.1}$$

Orthonormal bases are particular examples of frames. Frames allow for a (stable) series expansion of any $f \in \mathcal{K}$ of the form

$$f \;=\; Fu \;:=\; \sum_{\lambda\in\Lambda} u_\lambda \psi_\lambda \tag{2.2}$$

where $u = (u_\lambda)_{\lambda\in\Lambda} \in \ell_2(\Lambda)$. The linear operator $F : \ell_2(\Lambda) \to \mathcal{K}$ is called the *synthesis map* in frame theory. It is bounded due to the frame inequality (2.1). In contrast to orthonormal bases, the coefficients u_λ need not be unique, in general. For more information on frames we refer to [9].

By using frames the problem of recovering $f_\ell \in \mathcal{K}$ can be restated in terms of frame coefficients in $\ell_2(\Lambda, \mathbb{R}^M)$. To this end we introduce the operators $T_{\ell,j} = A_{\ell,j}F : \ell_2(\Lambda) \to \mathcal{H}_\ell$ and

$$T : \ell_2(\Lambda, \mathbb{R}^M) \to \mathcal{H}, \quad Tu \;=\; \left(\sum_{\ell=1}^{M} T_{\ell,j} u^\ell\right)_{j=1}^{N} \;=\; \left(\sum_{\ell=1}^{M} A_{\ell,j} F u^\ell\right)_{j=1}^{N},$$

where $\mathcal{H} := \bigoplus_{j=1}^N \mathcal{H}_j$ is the Hilbert space equipped with the inner product $\langle \sum_j g_j, \sum_j h_j \rangle := \sum_j \langle g_j, h_j \rangle$ with $g_j, h_j \in \mathcal{H}_j$. Then we want to solve approximatively the equation

$$g_j \;=\; \sum_{\ell=1}^{M} T_{\ell,j} u^\ell, \quad j = 1, \ldots, N.$$

Resuming the data vector into $g = (g_j)_{j=1,\ldots,M} \in \mathcal{H}$ the above equation can be written as

$$g \;=\; Tu. \tag{2.3}$$

Once the solution $u = (u_\lambda^\ell)$ is determined we obtain a reconstruction of the vectors of interest by means of $f_\ell = F u^\ell = \sum_\lambda u_\lambda^\ell \psi_\lambda$.

In practice, it often happens that the problem of recovering u from g is ill-posed or ill-conditioned, i.e., the operator T is not boundedly invertible or has very large condition number. A usual way out is regularization. In [24] we proposed to work with the following functional

$$J(u,v) \;=\; J_{\theta,\rho,\omega}^{(q)}(u,v) \;:=\; \|Tu - g|\mathcal{H}\|^2 + \sum_{\lambda\in\Lambda} v_\lambda \|u_\lambda\|_q + \sum_{\lambda\in\Lambda} \omega_\lambda \|u_\lambda\|_2^2 + \sum_{\lambda\in\Lambda} \theta_\lambda (\rho_\lambda - v_\lambda)^2, \tag{2.4}$$

where $q \in [1,\infty]$ and $\theta = (\theta_\lambda), \omega = (\omega_\lambda)$ and $\rho = (\rho_\lambda)$ are suitable sequences of positive parameters. The variable u is supposed to be in $\ell_2(\Lambda, \mathbb{R}^M)$ and $v \in \ell_{\infty,\rho^{-1}}(\Lambda)_+$, i.e., the subset of positive sequences in $\ell_{\infty,\rho^{-1}}(\Lambda)$. Observe, that u_λ is a vector in $\mathbb{R}^M$ while v_λ is just a positive scalar for all $\lambda \in \Lambda$.

We are interested in the joint minimizer (u^*, v^*) of this functional, and u^* is then considered as a regularized solution of (2.3). The variable v is an auxiliary variable that plays the role of an indicator of the sparsity pattern.

Regularization always implicitly models the solution u. In other words, one assumes preknowledge about u. Our functional (2.4) is related to *joint sparsity*. This means that we assume that all the components of the solution f_ℓ, $\ell = 1, \ldots, M$ can be well-represented as a linear combination of the *same* small subset of frame elements $\psi_\lambda, \lambda \in \Lambda_0$, where Λ_0 is a finite (small) subset of Λ, i.e.,

$$f_\ell \approx \sum_{\lambda\in\Lambda_0} u_\lambda^\ell \psi_\lambda.$$

It is important to note that Λ_0 does *not* depend on ℓ. For some intuition why the functional J promotes joint sparsity we refer to [24], but we will also see some reasons below. Further, we note that the functional is even interesting in the monochannel case $M = 1$. Then it just promotes usual sparsity and provides an alternative regularization to the one analyzed in [11].

At this point it is useful to denote the 'sparsity measure' by

$$\Phi^{(q)}(u,v) \;:=\; \Phi_{\theta,\rho,\omega}^{(q)}(u,v) \;:=\; \sum_{\lambda\in\Lambda} v_\lambda \|u_\lambda\|_q + \sum_{\lambda\in\Lambda} \omega_\lambda \|u_\lambda\|_2^2 + \sum_{\lambda\in\Lambda} \theta_\lambda (\rho_\lambda - v_\lambda)^2 \tag{2.5}$$

which allows to write

$$J(u,v) \;=\; \|Tu - g|\mathcal{H}\|^2 + \Phi^{(q)}(u,v).$$

In [24] we gave a criterion on the parameters ω, θ, ρ for the (strict) convexity of $\Phi^{(q)}(u,v)$, and hence of J, for the cases $q = 1, 2, \infty$. Let us a provide a slight generalization of this criterion.

Lemma 2.1. *Let $s_{\min} := \min \mathrm{Sp}(T^*T)$, where $\mathrm{Sp}(T^*T)$ denotes the spectrum of T^*T. A sufficient condition for the (strict) convexity of $J^{(q)}_{\theta,\rho,\omega}(u,v)$ is that the functions*

$$F_\lambda(x,y) \; := \; (\omega_\lambda + s_{\min})\|x\|_2^2 + y\|x\|_q + \theta_\lambda y^2, \quad x \in \mathbb{R}^M, y \geq 0$$

are (strictly) convex for all $\lambda \in \Lambda$. In the cases $q \in \{1, 2, \infty\}$ this is satisfied if

$$(\omega_\lambda + s_{\min})\theta_\lambda \geq \frac{\kappa_q}{4} \tag{2.6}$$

(with strict inequality for strict convexity), where

$$\kappa_q \; = \; \begin{cases} M, & q = 1 \\ 1, & q = 2, \\ 1, & q = \infty. \end{cases} \tag{2.7}$$

Proof. The discrepancy with respect to the data in the functional $J(u,v)$ can be written as

$$\begin{aligned} \|Tu - g|\mathcal{H}\|^2 \; &= \; \langle Tu, Tu\rangle - 2\langle Tu, g\rangle + \|g|\mathcal{H}\|^2 \; = \; \langle u, T^*Tu\rangle - 2\langle Tu, g\rangle + \|g|\mathcal{H}\|^2 \\ &= \; s_{\min}\|u\|_2^2 + \langle u, (T^*T - s_{\min}I)u\rangle - 2\langle Tu, g\rangle + \|g|\mathcal{H}\|^2, \end{aligned}$$

where I denotes the identity. Since $s_{\min} = \min \mathrm{Sp}(T^*T)$ the operator $T^*T - s_{\min}I$ is positive, and consequently the functional

$$u \mapsto \langle u, (T^*T - s_{\min}I)u\rangle - 2\langle Tu, g\rangle + \|g\|_2^2.$$

is convex. Thus, J is (strictly) convex if the functional

$$J'(u,v) \; = \; s_{\min}\|u\|_2^2 + \Phi^{(q)}(u,v) \; = \; \sum_{\lambda\in\Lambda} F_\lambda(u_\lambda, v_\lambda)$$

is (strictly) convex. Clearly, this is the case if and only if all the F_λ are (strictly) convex, which shows the first claim. The second claim for the cases $q = \{1, 2, \infty\}$ is shown precisely as in [24, Proposition 2.1]. ■

Usually one has $s_{\min} = 0$ and then (2.6) reduces to the condition already provided in [24]. However, there are cases where T^*T is invertible and then $s_{\min} > 0$, so (2.6) is weaker than $\omega_\lambda\theta_\lambda \geq \kappa_q/4$ in [24]. Further, we expect that condition (2.6) with suitable κ_q is also sufficient in the general case $q \in [1, \infty]$.

2.3 An algorithm for the minimization of J

In [24] we developed an iterative algorithm for computing the minimizer of $J(u,v)$. It consists of alternating a minimization with respect to u and v. More formally, for some initial choice $v^{(0)}$, for example $v^{(0)} = (\rho_\lambda)_{\lambda\in\Lambda}$, we define

$$\begin{aligned} u^{(n)} &:= \arg\min_{u\in\ell_2(\Lambda,\mathbb{R}^M)} J(u, v^{(n-1)}), \\ v^{(n)} &:= \arg\min_{v\in\ell_{\infty,\rho^{-1}}(\Lambda)_+} J(u^{(n)}, v). \end{aligned} \tag{2.8}$$

The minimizer $v^{(n)}$ of $J(u^{(n)}, v)$ for fixed $u^{(n)}$ can be computed explicitly by the formula

$$v_\lambda^{(n)} = \begin{cases} \rho_\lambda - \frac{1}{2\theta_\lambda}\|u_\lambda^{(n)}\|_q, & \|u_\lambda^{(n)}\|_q < 2\theta_\lambda\rho_\lambda, \\ 0, & \text{otherwise} . \end{cases} \tag{2.9}$$

The minimization of $J(u, v^{(n-1)})$ with respect to u and fixed $v^{(n-1)}$ can be done by an thresholded Landweber iteration scheme similar to the one analyzed in [11]. So let $v = (v_\lambda)_{\lambda\in\Lambda}$ be a fixed positive sequence and $u_{(0)} \in \ell_2(\Lambda, \mathbb{R}^M)$ be some arbitrary initial point and define

$$u_{(m)} := U_{v,\omega}^{(q)}(u_{(m-1)} + T^*(g - Tu_{(m-1)})), \quad m \geq 1, \tag{2.10}$$

where

$$(U_{v,\omega}^{(q)}(u))_\lambda = (1 + \omega_\lambda)^{-1} S_{v_\lambda}^{(q)}(u_\lambda) \tag{2.11}$$

and

$$S_v^{(q)}(x) = x - P_{v/2}^{q'}(x), \quad x \in \mathbb{R}^M, \tag{2.12}$$

with $P_{v/2}^{q'}$ denoting the orthogonal projection onto the unit ball of radius $v/2$ in $\mathbb{R}^M$ with respect to the q'-norm where $1/q' + 1/q = 1$. For $q \in \{1, 2, \infty\}$ explicit formulas for $S_{v/2}^q$ are given in [24].

By extending the arguments in [11] it was shown in [24, Proposition 4.9] that the iteration (2.10) strongly converges to the minimizer of $K(u) = J(u, v)$ under mild conditions on v and ω.

In [24] we showed that algorithm (2.8) indeed computes the minimizer of the functional J. For technical reasons we assume that $\|T\| < 1$. This can always be achieved by a suitable scaling of the functional.

Theorem 2.2. *Let $1 \leq q \leq \infty$ and assume that J is strictly convex (see also Lemma 2.1). Moreover, we assume that $\ell_{2,\omega^{1/2}}(\Lambda, \mathbb{R}^M)$ is embedded into $\ell_2(\Lambda, \mathbb{R}^M)$, i.e., $\inf_{\lambda\in\Lambda}\omega_\lambda > 0$. Then the sequence $(u^{(n)}, v^{(n)})_{n\in\mathbb{N}}$ converges to the unique minimizer $(u^*, v^*) \in \ell_2(\Lambda, \mathbb{R}^M) \times \ell_{\infty,\rho^{-1}}(\Lambda)_+$ of J. The convergence of $u^{(n)}$ is weak in $\ell_2(\Lambda, \mathbb{R}^M)$ and that of $v^{(n)}$ holds componentwise.*
For the most interesting cases $q \in \{1, 2, \infty\}$, assume in addition

$$4\theta_\lambda\omega_\lambda \geq \sigma > \kappa_q \tag{2.13}$$

for all $\lambda \in \Lambda$, with κ_q in (2.7). Then the convergence of $u^{(n)}$ to u^* is also strong in $\ell_2(\Lambda, \mathbb{R}^M)$. Moreover, $v^{(n)} - v^*$ converges to 0 strongly in $\ell_{2,\theta}(\Lambda)$.

We note that condition (2.13) can actually be relaxed to

$$\inf_{\lambda\in\Lambda} 4(\omega_\lambda + s_{\min})\theta_\lambda > \kappa_q$$

in order to ensure strong convergence. A careful inspection of the proof of [24, Theorem 3.1] allows to relax also the condition $\inf_{\lambda\in\Lambda}\omega_\lambda > 0$ to $s_{\min} + \inf_{\lambda\in\Lambda}\omega_\lambda > 0$. Indeed this condition is sufficient to ensure the coerciveness (see Definition 1 below) of the functional and hence the existence of minimizers (see the direct method of the calculus of variation, e.g. [10, Theorem 1.15]). More details on the analysis of the algorithm and an implementable version can be found in the original paper [24].

3 Relation to Hard and Firm Thresholding

The functional $J = J^{(q)}_{\theta,\rho,\omega}$ depends on many parameters. So far their role was not yet completely clarified. It turns out that there is an intriguing relationship to hard-thresholding, which explains the parameters as well.

3.1 A simple monochannel case

For the sake of simple illustration we start with the monochannel case $M = 1$ and the parameter $\omega = 0$ for the moment. (The choice of q becomes clearly irrelevant if $M = 1$). Further, we assume that $(\psi_\lambda)_\lambda$ is actually an orthonormal basis of $\mathcal{H}$ and $A : \mathcal{H} \to \mathcal{H}$ is unitary, so that T^*T is the identity operator on $\ell_2(\Lambda)$ and T^* is an isometry. Then

$$\|Tu - g|\mathcal{H}\|^2 = \|u - T^*g\|_2^2. \tag{3.1}$$

Setting $f = T^*g \in \ell_2(\Lambda)$ we consequently study the functional

$$\begin{aligned} J(u,v) \ = \ J_{\theta,\rho}(u,v) \ &= \ \|Tu - g\|_2^2 + \sum_{\lambda\in\Lambda} v_\lambda |u_\lambda| + \sum_{\lambda\in\Lambda} \theta_\lambda(\rho_\lambda - v_\lambda) \\ &= \sum_{\lambda\in\Lambda} \left[(u_\lambda - f_\lambda)^2 + v_\lambda|u_\lambda| + \theta_\lambda(\rho_\lambda - v_\lambda)^2 \right]. \end{aligned}$$

Due to Lemma 2.1 and since $s_{\min}(T^*T) = s_{\min}(I) = 1$ a sufficient (and actually necessary) condition for convexity of J is

$$\theta_\lambda \geq 1/4 \qquad \text{for all } \lambda \in \Lambda,$$

and J is strictly convex in case of a strict inequality. In our special case, J decouples as the sum

$$J(u,v) \ = \ \sum_{\lambda\in\Lambda} G_{\theta_\lambda,\rho_\lambda;f_\lambda}(u_\lambda, v_\lambda)$$

where

$$G_{\theta,\rho;z}(x,y) \ = \ (x-z)^2 + y|x| + \theta(\rho - y)^2, \qquad x \in \mathbb{R}, y \geq 0.$$

Hence, the component $(u^*_\lambda, v^*_\lambda)$, $\lambda \in \Lambda$, of the minimizer (u^*, v^*) of $J(u,v)$ is the minimizer of $G_{\theta_\lambda,\rho_\lambda;f_\lambda}(x,y)$.

Lemma 3.1. *Let $\rho > 0$, $\theta \geq 1/4$ and $z \in \mathbb{R}$. Then the minimizer (x^*, y^*) of $G_{\theta,\rho;z}(x,y)$ for $x \in \mathbb{R}, y \geq 0$ is given by*

$$\begin{aligned} x^* \ &= \ h_{\theta,\rho}(z) \\ y^* \ &= \ \begin{cases} \rho - \frac{1}{2\theta}|x^*|, & |x^*| < 2\theta\rho, \\ 0, & \textit{otherwise}, \end{cases} \end{aligned}$$

where

$$h_{\theta,\rho}(z) \ = \ \begin{cases} 0, & |z| \leq \rho/2, \\ \frac{4\theta}{4\theta-1}\left(z - \operatorname{sign}(z)\frac{\rho}{2}\right), & \rho/2 < |z| \leq 2\theta\rho, \\ z, & |z| > 2\theta\rho. \end{cases} \tag{3.2}$$

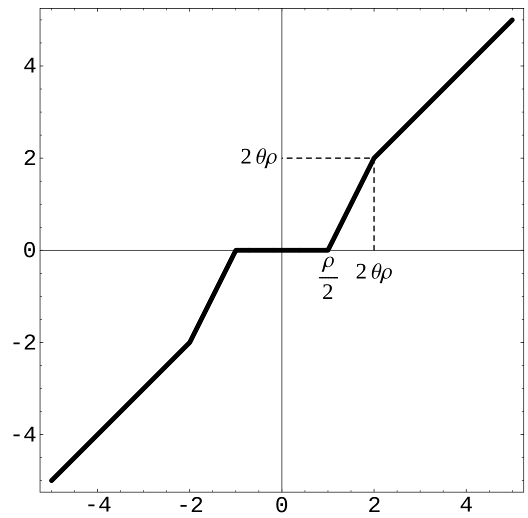

Figure 1: Typical shape of the function $h_{\theta,\rho}$. Here the parameters are $\rho = 2$ and $\theta = 1/2$.

Proof. The statement follows from a straightforward computation, but can also be deduced as a special case of Theorem 3.2 below (considering $\omega = 0$ and $q = 2$ for instance). ∎

Note that for $\theta = 1/4$ the function $h_{1/4,\rho}$ equals the hard thresholding function,

$$h_{1/4,\rho}(z) = h_\rho(z) := \begin{cases} 0, & |z| \leq \frac{\rho}{2} \\ z, & |z| > \frac{\rho}{2}. \end{cases}$$

In particular, hard-thresholding can be interpreted in terms of the (joint) minimizer of the functional

$$J(u,v) \;=\; \|u-f\|_2^2 + \sum_{\lambda\in\Lambda} v_\lambda |u_\lambda| + \frac{1}{4}\sum_{\lambda\in\Lambda}(\rho_\lambda - v_\lambda)^2,$$

and the minimizer is even unique although the functional is convex but not strictly convex. Note that it can be shown directly that also for $\theta < 1/4$ the minimizer of the functional J is still unique and coincides with the one for $\theta = 1/4$, although the functional is then even not convex any more.

Hence, not only soft-thresholding, but also hard-thresholding is related to the minimizer of a certain convex functional. This observation applies for instance to wavelet thresholding.

In the case $\theta > 1/4$ the function $h_{\theta,\rho}$ is the *firm thresholding* operator introduced in [25], see Figure 1 for a plot.

Furthermore, letting $\theta \to \infty$ in the above lemma, we recover the soft-thresholding function,

$$\lim_{\theta\to\infty} h_{\theta,\rho}(z) \;=\; s_\rho(z) \;=\; \begin{cases} 0, & |z| \leq \frac{\rho}{2} \\ z - \operatorname{sign}(z)\frac{\rho}{2}, & |z| > \frac{\rho}{2}. \end{cases}$$

Hence, $h_{\theta,\rho}$ can be interpreted as an interpolation between soft and hard thresholding.

3.2 The multichannel case with identity operator

Let us now consider the general multichannel case $M \geq 1$ with non-trivial parameter ω_λ. However, we again assume that T is as simple as in the previous section. In light of (3.1) we actually take T to be the identity on $\ell_2(\Lambda)$.

Our functional now has the form

$$J(u,v) \;=\; J^{(q)}_{\theta,\rho,\omega}(u,v) \;=\; \|u-f\|_2^2 + \sum_{\lambda\in\Lambda} \omega_\lambda \|u_\lambda\|_2^2 + \sum_{\lambda\in\Lambda} v_\lambda \|u_\lambda\|_q + \sum_{\lambda\in\Lambda} \theta_\lambda(\rho_\lambda - v_\lambda)^2 \tag{3.3}$$

with $u \in \ell_2(\Lambda, \mathbb{R}^M)$ and $v \in \ell_{\infty,\rho^{-1}}(\Lambda)_+$. By Lemma 2.1 a sufficient (and actually necessary) condition for convexity of J in the cases $q \in \{1, 2, \infty\}$ is

$$(1+\omega_\lambda)\theta_\lambda \geq \frac{\kappa_q}{4}$$

with κ_q as in (2.7). The functional J decouples as the following sum,

$$J(u,v) \;=\; \sum_{\lambda\in\Lambda} G^{(q)}_{\theta_\lambda,\rho_\lambda,\omega_\lambda;f_\lambda}(u_\lambda, v_\lambda)$$

with

$$G^{(q)}_{\theta,\rho,\omega;z}(x,y) \;:=\; \|x-z\|_2^2 + \omega\|x\|_2^2 + y\|x\|_q + \theta(\rho-y)^2, \quad x \in \mathbb{R}^M, y \in \mathbb{R}_+. \tag{3.4}$$

As in the previous section the minimization of J reduces to determining the minimizer of the function $G^{(q)}_{\theta,\rho,\omega;z}$ on $\mathbb{R}^M \times \mathbb{R}_+$.

Before stating the theoretical result let us introduce the following functions for $q = 1, 2, \infty$, respectively. For $q = 2$, $\theta > 1/4$ and $z \in \mathbb{R}^M$ we define

$$h^{(2)}_{\theta,\rho}(z) \;:=\; \begin{cases} 0, & \|z\|_2 \leq \rho/2, \\ \frac{4\theta}{4\theta-1}\frac{\|z\|_2-\rho/2}{\|z\|_2} z, & \rho/2 < \|z\|_2 \leq 2\theta\rho, \\ z, & \|z\|_2 > 2\theta\rho. \end{cases}$$

Now let $q = 1$, $\theta > M/4$ (ensuring strict convexity) and $z \in \mathbb{R}^M$. Then we distinguish different cases.

1. If $\|z\|_\infty < \rho/2$ then
$$h^{(1)}_{\theta,\rho}(z) \;:=\; 0.$$
2. If $\|z\|_1 \geq 2\theta\rho$ then
$$h^{(1)}_{\theta,\rho}(z) \;:=\; z.$$
3. If $\|z\|_\infty \geq \rho/2$ and $\|z\|_1 < 2\theta\rho$ then we order the entries of z by magnitude, $|z_{\ell_1}| \geq |z_{\ell_2}| \geq \ldots \geq |z_{\ell_M}|$. For $n = 1, \ldots, M$ define
$$t_n(z) \;:=\; \rho/2 - \frac{\sum_{j=1}^n |z_{\ell_j}| - \frac{n\rho}{2}}{4\theta - n}. \tag{3.5}$$

As follows from the proof of the next theorem there exists a unique $n \in \{1, \dots, M\}$ such that $\sum_{j=1}^{n} |z_{\ell_j}| \geq n\rho/2$,

$$|z_{\ell_n}| \geq t_n(z) \tag{3.6}$$

and

$$|z_{\ell_{n+1}}| < t_n(z) \tag{3.7}$$

(where the latter condition is void if $n = M$). With this particular n we define the components of $h^{(1)}_{\theta,\rho}(z)$ as

$$\begin{aligned} (h^{(1)}_{\theta,\rho}(z))_{\ell_j} &:= z_{\ell_j} - \operatorname{sign}(z_{\ell_j}) t_n(z), & j &= 1, \dots, n, \\ (h^{(1)}_{\theta,\rho}(z))_{\ell_j} &:= 0, & j &= n+1, \dots, M. \end{aligned}$$

Finally, let $q = \infty$, $\theta > 1/4$ and $z \in \mathbb{R}^M$. Again we have to distinguish several cases.

1. If $\|z\|_1 < \rho/2$ then
$$h^{(\infty)}_{\theta,\rho}(z) = 0.$$

2. If $\|z\|_\infty \geq 2\theta\rho$ then
$$h^{(\infty)}_{\theta,\rho}(z) = z.$$

3. If $\|z\|_1 \geq \rho/2$ and $\|z\|_\infty < 2\theta\rho$ then we order the coefficients of z by magnitude, $|z_{\ell_1}| \geq |z_{\ell_2}| \geq \ldots \geq |z_{\ell_M}|$. Define
$$s_n(z) := \frac{4\theta}{4\theta n - 1} \left(\sum_{j=1}^{n} |z_{\ell_j}| - \rho/2 \right).$$
Let m be the minimal number in $\{1, \dots, M\}$ such that $s_m(z) \geq 0$. (Such m exists since $s_M(z) \geq 0$ follows from $\|z\|_1 \geq \rho/2$.) As follows from the proof of the next theorem there exists a unique $n \in \{m, \dots, M\}$ such that
$$|z_{\ell_n}| \geq s_{n-1}(z) \quad \text{and} \quad |z_{\ell_{n+1}}| < s_n(z)$$
(where the first condition is void if $n = 1$ and the second condition is void if $n = M$). Then we define the components of $h^{(\infty)}_{\theta,\rho}$ as
$$\begin{aligned} (h^{(\infty)}_{\theta,\rho}(z))_{\ell_j} &:= \operatorname{sign}(z_{\ell_j}) s_n(z), & j &= 1, \dots, n, \\ (h^{(\infty)}_{\theta,\rho}(z))_{\ell_j} &:= z_{\ell_j}, & j &= n+1, \dots, M. \end{aligned}$$

These functions $h^{(q)}_{\theta,\rho}$ provide different generalizations of the firm shrinkage function $h_{\theta,\rho}$ in (3.2) to the multichannel case. As shown in the next result they are intimately related to the minimizer of the function $G^{(q)}_{\theta,\rho,\omega;z}$.

Theorem 3.2. *Let $q \in \{1, 2, \infty\}$ and $z \in \mathbb{R}^M$. Assume*

$$(\omega + 1)\theta > \kappa_q/4 \tag{3.8}$$

with κ_q in (2.7) ensuring strict convexity of the function $G^{(q)}_{\theta,\rho,\omega;z}$ in (3.4) by Lemma 2.1. Then the minimizer $(u, v) \in \mathbb{R}^M \times \mathbb{R}_+$ of $G^{(q)}_{\theta,\rho,\omega;z}(x, y)$ over $(x, y) \in \mathbb{R}^M \times \mathbb{R}_+$ is given by

$$\begin{aligned} u &= (1+\omega)^{-1} h^{(q)}_{\theta(1+\omega),\rho}(z), \\ v &= \begin{cases} \rho - \frac{\|u\|_q}{2\theta}, & \|u\|_q < 2\theta\rho, \\ 0, & \textit{otherwise} . \end{cases} \end{aligned} \tag{3.9}$$

The proof of this theorem is rather long and technical, and therefore postponed to the Appendix. We note that condition (3.8) is required to ensure uniqueness of the minimizer of $G_{\theta,\rho,\omega;z}$. In case of equality in (3.8) a variant of the above theorem still holds. Only the uniqueness of n in the definition of the function $h^{(q)}_{\theta,\rho}$ for $q = 1$ and $q = \infty$ is not clear yet, but any valid n would yield a minimizer of $G_{\theta,\rho,\omega;z}$.

Now, the minimizer (u^*, v^*) of the functional J for trivial operator T in (3.3) is clearly given by

$$u^* = H^{(q)}_{\theta,\rho,\omega}(f), \tag{3.10}$$

$$v^* = V^{(q)}_{\theta,\rho}(u^*), \tag{3.11}$$

where

$$\left(H^{(q)}_{\theta,\rho,\omega}(f) \right)_\lambda := (1+\omega_\lambda)^{-1} h^{(q)}_{\theta_\lambda(1+\omega_\lambda),\rho_\lambda}(f_\lambda), \tag{3.12}$$

and

$$\left(V^{(q)}_{\theta,\rho}(u^*) \right)_\lambda := \begin{cases} \rho_\lambda - \frac{1}{2\theta_\lambda}\|u^*_\lambda\|_q, & \|u^*_\lambda\|_q < 2\theta_\lambda\rho_\lambda \\ 0, & \text{otherwise} . \end{cases} \tag{3.13}$$

We note the following relation to the damped soft-thresholding operator $U^{(q)}_{v,\omega}$ in (2.11), which will be useful later.

Lemma 3.3. *Suppose $(1 + \omega_\lambda)\theta_\lambda > \kappa_q/4$ for all $\lambda \in \Lambda$. Let $v = V^{(q)}_{\theta,\rho}(H_{\theta,\rho,\omega}(f))$. Then*

$$H^{(q)}_{\theta,\rho,\omega}(f) = U^{(q)}_{v,\omega}(f).$$

Proof. Let (u^*, v^*) be the minimizer of the functional J in (3.3). Then $u^* = H^{(q)}_{\theta,\rho,\omega}(f)$ and $v^* = V^{(q)}_{\theta,\rho}(u)$ by (3.10) and (3.11). Since (u^*, v^*) minimizes $J(u, v)$, we have in particular $u^* = \arg\min_u J(u, v^*)$. By Lemma 4.1 in [24] it holds $u^* = U^{(q)}_{v^*,\omega}(f)$, which shows the claim. ■

Finally note that there is also the following alternative iterative way of computing the functions $h_{\theta,\rho}$.

Proposition 3.4. *Let $q \in \{1, 2, \infty\}$ and $4\theta \geq \kappa_q$. For $z \in \mathbb{R}^M$ and some $v^{(0)} \in \mathbb{R}_+$ define for $n \geq 1$*

$$\begin{aligned} z^{(n)} &= S^{(q)}_{v^{(n-1)}}(z) \;=\; x - P^{q'}_{v^{(n-1)}/2}(x) \\ v^{(n)} &= \begin{cases} \rho - \frac{1}{2\theta}\|z^{(n)}\|_q, & \|z^{(n)}\|_q < 2\theta\rho \\ 0, & \text{otherwise} . \end{cases} \end{aligned}$$

Then $z^{(n)}$ converges and $\lim_{n\to\infty} z^{(n)} = h^{(q)}_{\theta,\rho}(z)$. Moreover, if $4\theta > \kappa_q$ then we have the error estimate $|z^{(n)} - h^{(q)}_{\theta,\rho}(z)| \leq \gamma |z^{(n-1)} - h^{(q)}_{\theta,\rho}(z)|$ with $\gamma := \frac{\kappa_q}{4\theta} < 1$.

Proof. By Lemma 2.1 the corresponding function

$$J_z(u, v) \;=\; \|u - z\|_2 + v\|u\|_q + \theta(\rho - v)^2, \quad u \in \mathbb{R}^M, v \in \mathbb{R}_+, \tag{3.14}$$

is convex. The proposed iteration corresponds precisely to the scheme (2.8) and by Theorem 4.3 the scheme thus converges. The error estimate follows from Proposition 5.4 in [24]. ∎

Convergence of the scheme in the previous lemma holds even for general $q \in [1, \infty]$ provided the parameters are such that the corresponding functional in (3.14) is convex (although it is not completely clear yet that also the corresponding error estimate is true). However, it remains open whether a practical way of computing the projection $P^{q'}_{v/2}$ exists for values of q different from $1, 2, \infty$.

4 Iterative Thresholding Algorithms

Now we return to the analysis of the functional J with a general bounded operator T and $M \geq 1$ channels. By rescaling J we may assume without loss of generality that $\|T\| < 1$. However, note that rescaling changes the parameters θ, ω and $s_{\min} = s_{\min}(T^*T)$, so that eventually one has to take care not to destroy the convexity condition

$$\theta_\lambda(s_{\min}(T^*T) + \omega_\lambda) \geq \kappa_q/4. \tag{4.1}$$

We will now formulate and analyze a new algorithm for the minimization of J with non-trivial operator T. In contrast to the algorithm (2.8) analyzed in [24] it consists only of a single iteration scheme rather than a double minimization algorithm.

We first need to introduce surrogate functionals similar to the one in [11]. For some additional parameter $a \in \ell_2(\Lambda, \mathbb{R}^M)$ let

$$J^s(u, v; a) \;:=\; J(u, v) + \|u - a\|_2^2 - \|T(u - a)|\mathcal{H}\|^2.$$

Our iterative algorithm reads then as follows. For some arbitrary $u^{(0)} \in \ell_2(\Lambda, \mathbb{R}^M)$ we let

$$(u^{(n)}, v^{(n)}) \;:=\; \arg\min_{(u,v)} J^s(u, v; u^{(n-1)}), \quad n \geq 1. \tag{4.2}$$

The minimizer of $J^s(u, v; a)$ can be computed explicitly as we explain now. Denoting by $\Phi^{(q)}(u, v)$ the 'sparsity measure' defined in (2.5) a straightforward calculation yields

$$\begin{aligned} J^s(u, v; a) &= \|Tu - g|\mathcal{H}\|^2 - \|Tu - Ta|\mathcal{H}\|^2 + \|u - a\|_2^2 + \Phi^{(q)}(u, v) \\ &= \|u - (a + T^*(g - Ta))\|_2^2 + \Phi^{(q)}(u, v) + \|g|\mathcal{H}\|^2 - \|Ta|\mathcal{H}\|^2 + \|a\|_2^2 - \|a + T^*(g - Ta)\|_2^2. \end{aligned}$$

Since the terms after $\Phi^{(q)}(u, v)$ are constant with respect to u and v it follows that

$$\arg\min_{(u,v)} J^s(u, v; a) \;=\; \arg\min_{(u,v)} J'(u, v; a)$$

where

$$J'(u, v; a) \;=\; \|u - (a + T^*(g - Ta))\|_2^2 + \Phi^{(q)}(u, v).$$

We note that J' and, hence, $J^s(u, v; a)$ (for fixed a) is strictly convex if

$$\theta_\lambda(1 + \omega_\lambda) > \kappa_q/4$$

by Lemma 2.1. Since J' coincides with J where T is replaced by the identity and g by $a + T^*(g - Ta)$ we can invoke the results of the previous section to compute the minimizer (u^*, v^*) of J' and of $J^s(u, v; a)$. Indeed, if $q \in \{1, 2, \infty\}$ and $\theta_\lambda(1+\omega_\lambda) > \kappa_q/4$ for all $\lambda \in \Lambda$ then

$$u^* \;=\; H^{(q)}_{\theta,\rho,\omega}(a + T^*(g - Ta)), \tag{4.3}$$

and $v^* = V^{(q)}_{\theta,\rho}(u^*)$ with $H^{(q)}_{\theta,\rho,\omega}$ and $V^{(q)}_{\theta,\rho}$ defined in (3.12) and (3.13). It immediately follows that the algorithm in (4.2) reads

$$u^{(n)} \;=\; H^{(q)}_{\theta,\rho,\omega}(u^{(n-1)} + T^*(g - Tu^{(n-1)})). \tag{4.4}$$

It is actually not necessary to compute all the corresponding $v^{(n)}$'s. The final v^* can easily be computed by $v^* = V^{(q)}_{\theta,\rho}(u^*)$ if one is interested in it.

Algorithm (4.4) is a thresholded Landweber iteration. We note, however, that we cannot treat pure hard thresholding in this way, as this requires $\theta = 1/4$ and $\omega = 0$. Since $\|T\| < 1$ we have certainly also $s_{\min} = s_{\min}(T^*T) < 1$ and hence the convexity condition $\frac{1}{4}s_{\min} \geq \kappa_q/4$ cannot be satisfied. Moreover, if $s_{\min} = 0$ (which often happens in inverse problems) then we have to take $\omega_\lambda > 0$, which enforces a damping in the thresholding operation. Nevertheless, an "interpolation" between soft and hard thresholding is possible.

Before investigating the convergence of the thresholding algorithm (4.4) let us state an immediate implication of the previous achievements.

Proposition 4.1. *If $\|T\| < 1$ and $4(1+\omega_\lambda)\theta_\lambda > \kappa_q$ for all $\lambda \in \Lambda$ (ensuring strict convexity of the surrogate functional J^s) then a minimizer (u^*, v^*) of J satisfies the fixed point relation*

$$\begin{aligned} u^* &= H^{(q)}_{\theta,\omega,\rho}(u^* + T^*(g - Tu^*)),\\ v^* &= V^{(q)}_{\theta,\rho}(u^*). \end{aligned}$$

Conversely, if J is convex and (u^*, v^*) satisfies the above fixed point equation then it is a minimizer of J.

Proof. Observe that $J^s(u^*, v^*; u^*) = J^s(u^*, v^*)$, but in general $J^s(u, v; a) \geq J(u, v)$ for all (u, v) because $\|T\| < 1$. Hence, if (u^*, v^*) minimizes $J(u, v)$ then it also minimizes $J^s(u, v; u^*)$ and by (4.3) (noting that $4(1 + \omega_\lambda)\theta_\lambda > \kappa_q$) the stated fixed point equation is satisfied.

Conversely, if (u^*, v^*) satisfies the fixed point equation then by Theorem 3.2 $(u^*_\lambda, v^*_\lambda)$ is the minimizer of $G_\lambda = G_{\theta_\lambda,\rho_\lambda,\omega_\lambda;z}$ for $z = (u^* + T^*(g - Tu^*))_\lambda$, i.e., 0 is contained in the subdifferential of G_λ for all $\lambda \in \Lambda$. If J is convex then by Proposition 3.5 in [24] the subdifferential of J at (u, v) contains the set

$$DJ(u,v) \;=\; (2T^*(Tu-g), 0) + D\Phi^{(q)}(u,v),$$

where

$$D\Phi^{(q)}(u,v) \;=\; \{(\xi,\eta) \in \ell_2(\Lambda,\mathbb{R}^M) \times \ell_{1,\rho}(\Lambda), \xi_\lambda \in v_\lambda \partial\|\cdot\|_q(u_\lambda) + 2\omega_\lambda u_\lambda,$$
$$\eta_\lambda \in \|u_\lambda\|_q \partial s^+(v_\lambda) + 2\theta_\lambda(v_\lambda - \rho_\lambda),\;\; \lambda \in \Lambda\}$$

where $\partial s^+(x) = \{1\}$ for $x > 1$ and $\partial s^+(0) = (-\infty, 1]$. Using Lemma A.1 it is then straightforward to verify that 0 is contained in $DJ(u^*, v^*) \subset \partial J(u^*, v^*)$, and hence, (u^*, v^*) minimizes J. ∎

Note that the first part of the above theorem does not require convexity of J as the general convexity condition $(s_{\min}(T^*T) + \omega_\lambda)\theta_\lambda \geq \kappa_q/4$ is stronger than the required condition since $s_{\min} < 1$.

For later reference we note that the minimizer of J actually satisfies also another fixpoint relation in terms of the soft-thresholding operator:

Proposition 4.2. *If $\|T\| < 1$ then a minimizer (u^*, v^*) of J satisfies the fixed point equations*

$$u^* \;=\; U^{(q)}_{v^*,\omega}(u^* + T^*(g - T^*a)),$$
$$v^* \;=\; V^{(q)}_{\theta,\rho}(u^*),$$

with $U^{(q)}_{v,\omega}$ defined by (2.11).

Proof. The relation $v^* = V^{(q)}_{\theta,\rho}(u^*)$ is clear. Similarly as in the previous proof we have

$$J(u^*,v^*) = \min_u J(u,v^*) \;=\; \min_u J^s(u,v^*;u^*),$$

and u^* minimizes $J^s(u, v^*; u^*)$ for fixed v^* and u^*. By Lemma 4.1 in [24] it follows that $u^* = U^{(q)}_{v^*,\omega}(u^* + T^*(g - Tu^*))$ as claimed. ∎

Note that the previous result does not pose any restrictions on the parameters θ, ρ, ω. In particular, $J(u, v)$ may even fail to be jointly convex in u, v. Furthermore, the two relations in Theorem 4.2 are coupled whereas the first relation in Theorem 4.1 is independent of the second one.

4.1 Convergence of the iterative algorithm

Let us now investigate the convergence of the iterative algorithm (4.4).

Theorem 4.3. *Let $q \in \{1, 2, \infty\}$ and assume that $\|T\| < 1$ and*

$$\inf_{\lambda\in\Lambda} 4\theta_\lambda(s_{\min} + \omega_\lambda) > \kappa_q \tag{4.5}$$

with $s_{\min} = \min \operatorname{Sp}(T^*T)$ (ensuring strict convexity of J by Lemma 2.1). Then for any choice $u^{(0)} \in \ell_2(\Lambda; \mathbb{R}^M)$ the iterative algorithm (4.2), i.e.,

$$u^{(n)} := H^{(q)}_{\theta,\rho,\omega}\left(u^{(n-1)} + T^*(g - Tu^{(n-1)})\right), \tag{4.6}$$

converges strongly to a fixed point $u^* \in \ell_2(\Lambda; \mathbb{R}^M)$ and the couple (u^*, v^*) with $v^* = V^{(q)}_{\theta,\rho}(u^*)$ is the unique minimizer of J. Moreover, we have the error estimate

$$\|u^{(n)} - u^*\|_2 \le \beta^n \|u^{(0)} - u^*\|_2 \tag{4.7}$$

with $\beta := \sup_{\lambda\in\Lambda} \frac{4\theta_\lambda(1-s_{\min})}{4\theta_\lambda(1+\omega_\lambda)-\kappa_q} < 1$.

An essential ingredient for the proof of this theorem is the following.

Lemma 4.4. *Assume $q \in \{1, 2, \infty\}$ and $4\theta_\lambda(1+\omega_\lambda) > \kappa_q$ for all $\lambda \in \Lambda$. Then the operators $H^{(q)}_{\theta,\rho,\omega}$ are Lipschitz continuous,*

$$\|H^{(q)}_{\theta,\rho,\omega}(y) - H^{(q)}_{\theta,\rho,\omega}(z)\|_2 \le L\|y - z\|_2$$

with constant $L := \sup_{\lambda\in\Lambda} \frac{4\theta_\lambda}{4\theta_\lambda(1+\omega_\lambda)-\kappa_q}$.

Proof. By Lemma 3.3 we have $H^{(q)}_{\theta,\rho,\omega}(z) = U^{(q)}_{v,\omega}(z)$ with $v = V^{(q)}_{\theta,\rho}(H^{(q)}_{\theta,\rho,\omega}(z)) =: v(z)$. By the triangle inequality

$$\begin{aligned}
&\|(H^{(q)}_{\theta,\rho,\omega}(y))_\lambda - (H^{(q)}_{\theta,\rho,\omega}(z))_\lambda\|_2 \\
&\le \|U^{(q)}_{v(y),\omega}(y)_\lambda - U^{(q)}_{v(y),\omega}(z)_\lambda\|_2 + \|U^{(q)}_{v(y),\omega}(z)_\lambda - U^{(q)}_{v(z)),\omega}(z)_\lambda\|_2 \\
&= (1+\omega_\lambda)^{-1}\left[\|S^{(q)}_{v_\lambda(y)}(y_\lambda) - S^{(q)}_{v_\lambda(y)}(z_\lambda)\|_2 + \|S^{(q)}_{v_\lambda(y)}(z_\lambda) - S^{(q)}_{v_\lambda(z)}(z_\lambda)\|_2\right].
\end{aligned} \tag{4.8}$$

Since $S^{(q)}_{v_\lambda}(x) = x - P^{q'}_{v_\lambda/2}(x)$, where $P^{q'}_{v_\lambda/2}$ is the orthogonal projection onto the $\ell_{q'}$-ball of radius $v_\lambda/2$ the first term can be estimated by

$$\|S^{(q)}_{v_\lambda(y)}(y_\lambda) - S^{(q)}_{v_\lambda(y)}(z_\lambda)\|_2 \le \|y_\lambda - z_\lambda\|_2.$$

Further, it was proved in [24, Lemma 5.2] that $\|P^{q'}_v(x) - P^{q'}_w(x)\| \le K_q|v-w|$ for all $v, w \ge 0$, and $x \in \mathbb{R}^M$, with $K_1 = \sqrt{M}$ and $K_2 = K_\infty = 1$. The second term in (4.8) can thus be estimated by

$$\begin{aligned}
\|S^{(q)}_{v_\lambda(y)}(z_\lambda) - S^{(q)}_{v_\lambda(z)}(z_\lambda)\|_2 &= \|P^{q'}_{v_\lambda(y)/2}(z_\lambda) - P^{q'}_{v_\lambda(z)/2}(z_\lambda)\|_2 \le \frac{K_q}{2}|v_\lambda(y) - v_\lambda(z)| \\
&= \frac{K_q}{2}|V^{(q)}_{\theta,\rho}(H^{(q)}_{\theta,\rho,\omega}(y))_\lambda - V^{(q)}_{\theta,\rho}(H_{\theta,\rho,\omega}(z))_\lambda|.
\end{aligned}$$

Using the definition of $V_{\theta,\rho}^{(q)}$ in (3.13) and distinguishing different cases we obtain

$$\begin{aligned}|V_{\theta,\rho}^{(q)}(H_{\theta,\rho,\omega}^{(q)}(y))_\lambda - V_{\theta,\rho}^{(q)}(H_{\theta,\rho,\omega}^{(q)}(z))_\lambda| \;&\le\; \frac{1}{2\theta_\lambda}\Big|\|(H_{\theta,\rho,\omega}^{(q)}(y))_\lambda\|_q - \|(H_{\theta,\rho,\omega}^{(q)}(z))_\lambda\|_q\Big| \\ \le\; \frac{1}{2\theta_\lambda}\|(H_{\theta,\rho,\omega}^{(q)}(y))_\lambda - (H_{\theta,\rho,\omega}^{(q)}(z))_\lambda\|_q \;&\le\; \frac{R_q}{2\theta_\lambda}\|(H_{\theta,\rho,\omega}^{(q)}(y))_\lambda - (H_{\theta,\rho,\omega}^{(q)}(z))_\lambda\|_2,\end{aligned}$$

where $R_q = 1$ for $q \in \{2, \infty\}$ and $R_1 = \sqrt{M}$. Altogether we deduced

$$\begin{aligned}&\|(H_{\theta,\rho,\omega}^{(q)}(y))_\lambda - (H_{\theta,\rho,\omega}^{(q)}(z))_\lambda\|_2 \\ &\le\; (1+\omega_\lambda)^{-1}\left[\|y_\lambda - z_\lambda\|_2 + \frac{K_q R_q}{4\theta_\lambda}\|(H_{\theta,\rho,\omega}^{(q)}(y))_\lambda - (H_{\theta,\rho,\omega}^{(q)}(z))_\lambda\|_2\right].\end{aligned}$$

Noting that $K_q R_q = \kappa_q$ we obtain

$$\left(1 - \frac{\kappa_q}{4\theta_\lambda(1+\omega_\lambda)}\right)\|(H_{\theta,\rho,\omega}^{(q)}(y))_\lambda - (H_{\theta,\rho,\omega}^{(q)}(z))_\lambda\|_2 \;\le\; (1+\omega_\lambda)^{-1}\|y_\lambda - z_\lambda\|_2.$$

Summing over $\lambda \in \Lambda$ we finally obtain

$$\|H_{\theta,\rho,\omega}^{(q)}(y) - H_{\theta,\rho,\omega}^{(q)}(z)\|_2 \;\le\; \sup_{\lambda\in\Lambda}\frac{1}{(1+\omega_\lambda) - \frac{\kappa_q}{4\theta_\lambda}}\|y-z\|_2 \;=\; L\|y-z\|_2,$$

and the proof is completed. ∎

Proof of Theorem 4.3. Let Γ denote the operator

$$\Gamma(u) \;:=\; H_{\theta,\rho,\omega}^{(q)}(u + T^*(g - Tu)). \tag{4.9}$$

Then clearly, $u^{(n)} = \Gamma(u^{(n-1)})$. By Lemma 4.4 Γ is Lipschitz,

$$\begin{aligned}&\|\Gamma(y) - \Gamma(z)\|_2 \le \|H_{\theta,\rho,\omega}^{(q)}(y + T^*(g - Ty)) - H_{\theta,\rho,\omega}^{(q)}(z + T^*(g - Tz))\|_2 \\ &\le\; L\|y + T^*(g-Ty) - z - T^*(g - Tz)\|_2 \;=\; L\|(I - T^*T)(y-z)\|_2 \\ &\le L\|I - T^*T\|\,\|y-z\|_2 \;=\; L(1 - s_{\min})\|y-z\|_2 \;=\; \sup_{\lambda\in\Lambda}\frac{4\theta_\lambda(1-s_{\min})}{4\theta_\lambda(1+\omega_\lambda) - \kappa_q}\|y-z\|_2 \\ &=\; \beta\|y-z\|_2.\end{aligned} \tag{4.10}$$

Since by assumption $\beta < 1$ it follows from Banach's fixed point theorem that $u^{(n)}$ converges to the unique fixed point u^* of Γ and

$$\|u^{(n)} - u^*\|_2 = \|\Gamma(u^{(n-1)}) - \Gamma(u^*)\|_2 \le \beta\|u^{(n-1)} - u^*\|_2.$$

By induction we deduce (4.7). By Theorem 4.1 (u^*, v^*) with $v^* = V_{\theta,\rho}^{(q)}(u^*)$ is the unique minimizer of J. ∎

4.2 Convergence for the non-contractive case

In the previous section we have established with relatively simple arguments the convergence of the algorithm (4.2) to the minimizer of J under the condition (4.5). Thus, we provided an alternative to the first established algorithm (2.8) in [24].

It is intriguing to investigate what happens in the case when condition (4.5) is relaxed to

$$\inf_{\lambda \in \Lambda} 4\theta_\lambda (s_{\min} + \omega_\lambda) \geq \kappa_q, \tag{4.11}$$

and clearly, the remaining case is when there is equality in the above condition. The latter means that the iteration map Γ, see (4.9), is still non-expansive, but not strictly contractive anymore.

We will use Opial's fixed point theorem:

Theorem 4.5. *Let the mapping Γ from $\mathcal{H}$ to $\mathcal{H}$ satisfy the following conditions:*

(i) Γ is nonexpansive: for all $z, z' \in \mathcal{H}$, $\|\Gamma z - \Gamma z'\| \leq \|z - z'\|$;

(ii) Γ is asymptotically regular: for all $z \in \mathcal{H}$, $\|\Gamma^{n+1} z - \Gamma^n z\| \to 0$, for $n \to \infty$;

(iii) the set $\mathcal{F}$ of fixed points of Γ in $\mathcal{H}$ is not empty.

Then for all $z \in \mathcal{H}$, the sequence $(\Gamma^n z)_{n \in \mathbb{N}}$ converges weakly to a fixed point in $\mathcal{F}$.

A simple proof of this theorem can be found in [11].

The goal of our analysis in this section is to show that the map Γ defined in (4.9) and ruling the iteration of the algorithm (4.2) fulfills the requirements of Theorem 4.5. By condition (4.11) and (4.10) we have

$$\|\Gamma(z) - \Gamma(z')\|_2 \leq \|z - z'\|_2,$$

which shows the non-expansiveness (i). If $s_{\min} + \omega_\lambda \geq \gamma > 0$ for all $\lambda \in \Lambda$, then there exists a minimizer of $J(u, v)$ by coerciveness as already noted before. From Theorem 4.1 we have that such minimizers are fixed points of Γ, and hence also (iii) is satisfied. It remains to show (ii).

Lemma 4.6. *If $\|T\| < 1$ and $4(1 + \omega_\lambda)\theta_\lambda > \kappa_q$ for all $\lambda \in \Lambda$ then the mapping Γ is asymptotically regular.*

Proof. We first show that both $(J(u^{(n)}, V_{\theta,\rho}^{(q)}(u^{(n)})))_{n \in \mathbb{N}}$ and $(J^s(u^{(n+1)}, V_{\theta,\rho}^{(q)}(u^{(n+1)}); u^{(n)}))_{n \in \mathbb{N}}$ are nondecreasing sequences. By definition of $u^{(n+1)}$, we have

$$\begin{aligned} J(u^{(n+1)}, V_{\theta,\rho}^{(q)}(u^{(n+1)})) &\leq J^s(u^{(n+1)}, V_{\theta,\rho}^{(q)}(u^{(n+1)}); u^{(n)}) \leq J^s(u^{(n)}, V_{\theta,\rho}^{(q)}(u^{(n)}); u^{(n)}) \\ &= J(u^{(n)}, V_{\theta,\rho}^{(q)}(u^{(n)})), \end{aligned}$$

and

$$J^s(u^{(n+2)}, V_{\theta,\rho}^{(q)}(u^{(n+2)}); u^{(n+1)}) \leq J(u^{(n+1)}, V_{\theta,\rho}^{(q)}(u^{(n+1)})) \leq J^s(u^{(n+1)}, V_{\theta,\rho}^{(q)}(u^{(n+1)}); u^{(n)}).$$

Moreover, observe that

$$
\begin{aligned}
&J^s(u^{(n+1)}, V^{(q)}_{\theta,\rho}(u^{(n+1)}); u^{(n)}) - J^s(u^{(n+2)}, V^{(q)}_{\theta,\rho}(u^{(n+2)}); u^{(n+1)}) \\
&\geq J^s(u^{(n+1)}, V^{(q)}_{\theta,\rho}(u^{(n+1)}); u^{(n)}) - J(u^{(n+1)}, V^{(q)}_{\theta,\rho}(u^{(n+1)})) \\
&= \|u^{(n+1)} - u^{(n)}\|_2^2 - \|T(u^{(n+1)} - u^{(n)})|\mathcal{H}\|^2 \geq \underbrace{(1 - \|T\|^2)}_{:=C} \|u^{(n+1)} - u^{(n)}\|_2^2.
\end{aligned}
$$

Clearly, $C > 0$ and by the latter inequality we have

$$
\begin{aligned}
&C \sum_{n=0}^{N} \|u^{(n+1)} - u^{(n)}\|^2 \\
&\leq \sum_{n=0}^{N} \Big(J^s(u^{(n+1)}, V^{(q)}_{\theta,\rho}(u^{(n+1)}); u^{(n)}) - J^s(u^{(n+2)}, V^{(q)}_{\theta,\rho}(u^{(n+2)}); u^{(n+1)}) \Big) \\
&= \Big(J^s(u^{(1)}, V^{(q)}_{\theta,\rho}(u^{(1)}); u^{(0)}) - J^s(u^{(N+2)}, V^{(q)}_{\theta,\rho}(u^{(N+2)}); u^{(N+1)}) \Big) \\
&\leq J^s(u^{(1)}, V^{(q)}_{\theta,\rho}(u^{(1)}); u^{(0)}) < \infty.
\end{aligned}
$$

This implies $\sum_{n=0}^{\infty} \|u^{(n+1)} - u^{(n)}\|_2^2 < \infty$ and $\|u^{(n+1)} - u^{(n)}\| \to 0$ for $n \to \infty$. Therefore $\|\Gamma^{n+1}(u^{(0)}) - \Gamma^n(u^{(0)})\|_2 = \|u^{(n+1)} - u^{(n)}\|_2 \to 0$ and the mapping Γ is asymptotically regular. ■

By combining the previous achievements, we obtain the following convergence result.

Theorem 4.7. *If $\|T\| < 1$, $\inf_{\lambda\in\Lambda} 4\theta_\lambda(s_{\min} + \omega_\lambda) \geq \kappa_q$ and $s_{\min} + \inf_{\lambda\in\Lambda} \omega_\lambda > 0$ then for any initial choice $u^{(0)} \in \ell_2(\Lambda, \mathbb{R}^M)$, the sequence*

$$
u^{(n+1)} := H^{(q)}_{\theta,\rho,\omega} \Big(u^{(n)} + T^*(g - Tu^{(n)}) \Big), \tag{4.12}
$$

converges weakly to a fixed point u^* of Γ and $(u^*, V^{(q)}_{\theta,\rho}(u^*))$ is a minimizer of J. Moreover, if $4\theta_\lambda(s_{\min} + \omega_\lambda) > \kappa_q$ for all $\lambda \in \Lambda$ then u^* is the unique fixed point of Γ and $(u^*, V^{(q)}_{\theta,\rho}(u^*))$ is the unique minimizer of J.

We do not further insist in the task of investigating the strong convergence of the sequence $(u^{(n)})_{n\in\mathbb{N}}$ under the conditions of Theorem 4.7 when $\inf_{\lambda\in\Lambda} 4\theta_\lambda(s_{\min} + \omega_\lambda) = \kappa_q$. We refer to [11, Proposition 2.1, Proposition 3.10, Section 3.2] for results in this direction related to soft-thresholding.

5 On Variational Limits

In this section we investigate how the minimizers of $J = J^{(q)}_{\theta,\rho,\omega}$ vary when the parameters are changed.

5.1 Approaching soft-thresholding

We will now keep the sequence ρ fixed and let $\omega = \omega^{(k)}$ and $\theta = \theta^{(k)}$ vary with $k \in \mathbb{N}$. For brevity we denote the corresponding functionals by $J_{(k)} = J^{(q)}_{\theta^{(k)},\rho,\omega^{(k)}}$.

The result below reveals how one can continuously approach minimizers of the functional

$$K_\rho(u) := \|Tu - g|\mathcal{H}\|^2 + \sum_{\lambda\in\Lambda} \rho_\lambda \|u_\lambda\|_q,$$

by means of minimizers of $J_{(k)}$. K_ρ is closely related to the soft thresholding operator $S^{(q)}_\rho$ in (2.12), and its minimizer can be approximated by the algorithm (2.10) with $\omega_\lambda = 0$, which indeed is a pure soft-thresholded Landweber iteration, see [11, 24].

Theorem 5.1. *Let $q \in \{1, 2, \infty\}$. Suppose ρ is a sequence satisfying $\inf_{\lambda\in\Lambda} \rho_\lambda > 0$. Assume that the entries $\theta^{(k)}_\lambda$ are monotonically increasing with k for all λ and*

$$\lim_{k\to\infty} (\inf_{\lambda\in\Lambda} \theta^{(k)}_\lambda) = \infty. \tag{5.1}$$

Further suppose

$$\kappa_q < 4\omega^{(k)}_\lambda \theta^{(k)}_\lambda \le C \tag{5.2}$$

for some constant $C > \kappa_q$ and

$$\omega^{(k)}_\lambda - \frac{1}{4\kappa_{q'}\theta^{(k)}_\lambda} \le \omega^{(k-1)}_\lambda - \frac{1}{4\kappa_q \theta^{(k-1)}_\lambda} \tag{5.3}$$

for all $\lambda \in \Lambda$ and $k \in \mathbb{N}$, where q' denotes the dual index of q, i.e., $1/q' + 1/q = 1$ as usual. Denote by $(u^{(k)}, v^{(k)})$ the (unique) minimizer of $J_{(k)}(u, v) = J^{(q)}_{\theta^{(k)},\rho,\omega^{(k)}}(u, v)$. Then the accumulation points of the sequence $(u^{(k)})_{k\in\mathbb{N}}$ with respect to the weak topology in $\ell_2(\Lambda, \mathbb{R}^M)$ are minimizers of K_ρ. In particular, if the minimizer of K_ρ is unique then $u^{(k)}$ converges weakly to it.

The proof of this theorem uses some machinery from Γ-convergence [10] as a main tool. To state the corresponding result we first need to introduce some notion.

Definition 1. (a) A functional $F : X \to \overline{\mathbb{R}}$ on a topological space X satisfying the first axiom of countability (i.e., being metrizable) is called *lower semicontinuous* if for all x and all sequences x_k converging to x it holds $F(x) \le \liminf_k F(x_k)$.

(b) A function $F : X \to \overline{\mathbb{R}}$ is called coercive if for all $t \in \mathbb{R}$ the set $\{x : F(x) \le t\}$ is contained in a compact set.

The following well-known result can be achieved as a direct combination of [10, Proposition 5.7, Theorem 7.8, Corollary 7.20, Corollary 7.24]. For the sake of completeness we provide a proof, which implicitly uses techniques from Γ-convergence. For more details we refer to [10].

Theorem 5.2. *Let X be a topological space which satisfies the first axiom of countability. Assume that F_k, $k \in \mathbb{N}$, is a monotonically decreasing sequence of functionals on a topological space X that converges pointwise to a functional F, i.e., $F_{k+1}(x) \leq F_k(x)$ and $\lim_{k\to\infty} F_k(x) = F(x)$ for all $x \in X$. Assume that F is lower semicontinuous and coercive. Suppose that x_k minimizes F_k over X. Then the accumulation points of the sequence $(x_k)_{k\in\mathbb{N}}$ are minimizers of F. Moreover, if the minimizer of F is unique then x_k converges to it.*

Proof. Let (x_k) be a sequence converging to $x \in X$. By lower-semicontinuity and since F_k is monotonically decreasing we have

$$F(x) \leq \liminf_k F(x_k) \leq \liminf_k F_k(x_k). \tag{5.4}$$

Furthermore, by pointwise convergence we have

$$\inf\{\limsup_k F_k(x_k), x_k \to x\} \leq F(x)$$

where the infimum is taken over all sequences x_k converging x, and in the inequality it was used that this infimum is certainly less then the quantity attained for the constant sequence $x_k = x$. The above infimum is actually attained, see [10, Proposition 8.1 c) and d)], so there exists a sequence $x_k \to x$ such that $F(x) \geq \limsup_k F_k(x_k)$ and by (5.4) it follows $F(x) = \lim_k F_k(x_k)$, and furthermore $F(x) \geq \lim_k(\inf_{x\in X} F_k(x))$. Since x was arbitrary, it follows that

$$\inf_{x\in X} F(x) \geq \lim_k(\min_{x\in X} F_k(x)). \tag{5.5}$$

Since $F_k \geq F$, we have $\{x, F_k(x) \leq t\} \subset \{x, F(x) \leq t\}$ for all t and the latter is contained in a compact set by coerciveness. Thus, if x_k minimizes F_k for each k then the sequence x_k is contained in a compact set. Hence, we can extract a subsequence x_{k_j} which converges to one of the accumulation points x' of x_k. Then inequality (5.4) yields

$$\inf_{x\in X} F(x) \leq F(x') \leq \lim_j F_{k_j}(x_{k_j}) = \lim_{k_j}(\min_{x\in X} F_{k_j}(x)).$$

Together with (5.5) it follows that

$$F(x') = \inf_{x\in X} F(x) = \lim_j(\min_{x\in X} F_{k_j}(x)).$$

This means that x' minimizes F and we showed that all accumulation points of the sequence (x_k) are minimizers of F. Now if the minimizer of F is unique then with the same argument as above it follows that every subsequence of x_k contains another subsequence that converges to x'. But then x_k itself must converge to x'. ∎

Proof of Theorem 5.1. First we show that K_ρ is coercive and lower-semicontinuous with respect to the weak topology of $\ell_2(\Lambda, \mathbb{R}^M)$. Since $\inf_\lambda \rho_\lambda > 0$ we have

$$\|u\|_2 \leq \left(\sup_{\lambda\in\Lambda} \rho_\lambda^{-1}\right) \sum_{\lambda\in\Lambda} \rho_\lambda \|u_\lambda\|_2 \leq C_q \sum_{\lambda\in\Lambda} \rho_\lambda \|u_\lambda\|_q.$$

Hence, if u is such that $K_\rho(u) \le t$, then $\|u\|_2 \le C_q t$, which shows that $\{u \in \ell_2, K_\rho(u) \le t\}$ is contained in the ℓ_2 ball of radius $C_q t$, which is compact in the weak topology. Hence, K_ρ is coercive.

Since we are interested in minimization problems it suffices to consider our functionals on the set $X = \{u \in \ell_2, K_\rho(u) \le C\}$ for a sufficiently large C. Observe that by [10, Proposition 8.7] the space X is indeed metrizable with the weak topology inherited from $\ell_2(\Lambda, \mathbb{R}^M)$.

Now consider a sequence $(u^{(k)})$ which is weakly convergent to u. By weak convergence and lower semicontinuity of the $\mathcal{H}$ norm we have $\|Tu - g\|_{\mathcal{H}} \le \lim_k \|Tu^{(k)} - g\|_{\mathcal{H}}$. Weak convergence in ℓ_2 implies convergence of the components $u_\lambda^{(k)}$. Hence, by Fatou's lemma we further have

$$\sum_\lambda \rho_\lambda \|u_\lambda\|_q \;=\; \sum_\lambda \rho_\lambda \liminf_k \|u_\lambda^{(k)}\|_q \le \liminf_k \sum_\lambda \rho_\lambda \|u_\lambda^{(k)}\|_q.$$

This implies that K_ρ is lower-semicontinuous in X.

If $(u^{(k)}, v^{(k)})$ minimizes $J_{(k)}$ then $v^{(k)} = V^{(q)}_{\theta^{(k)},\rho^{(k)}}(u^{(k)})$. Hence, $(u^{(k)}, v^{(k)})$ is a minimizer of $J_{(k)}$ if and only if $u^{(k)}$ minimizes as well the functional

$$F_{(k)}(u) \;:=\; J_{(k)}(u, V^{(q)}_{\theta^{(k)},\rho^{(k)}}(u)).$$

Above we have already seen that the set X is bounded in the ℓ_2 norm, hence, if $u \in X$ then its components satisfy $\|u_\lambda\|_q \le C'$. By assumption (5.1) and since ρ_λ is bounded away from 0, there exists a $k_0 \in \mathbb{N}$ such that $\|u_\lambda\|_q \le 2\theta_\lambda^{(k)} \rho_\lambda$ for all $k \ge k_0$ and all $\lambda \in \Lambda$. Consequently

$$v_\lambda^{(k)} \;=\; \rho_\lambda - \frac{\|u_\lambda\|_q}{2\theta_\lambda^{(n)}}, \quad \forall \lambda \in \Lambda,$$

and the functional $F_{(k)}$ is given by

$$F_{(k)}(u) \;=\; \|Tu - g\|^2 + \sum_{\lambda\in\Lambda} \rho_\lambda \|u_\lambda\|_q + \sum_{\lambda\in\Lambda} \left(\omega_\lambda^{(k)} \|u_\lambda\|_2^2 - \frac{\|u_\lambda\|_q^2}{4\theta_\lambda^{(k)}} \right)$$

for all $k \ge k_0$ and $u \in X$. Clearly, it suffices to restrict all considerations to $k \ge k_0$.

Since the ℓ_q-norm on $\mathbb{R}^M$ is equivalent to the ℓ_2-norm it follows from (5.1) and (5.2) that

$$\lim_{k\to\infty} \sum_{\lambda\in\Lambda} \left(\omega_\lambda^{(k)} \|u_\lambda\|_2^2 - \frac{\|u_\lambda\|_q^2}{4\theta_\lambda^{(k)}} \right) = 0$$

for all $u \in X$. Hence $F_{(k)}$ converges pointwise to K_ρ on X. Further, note that (5.3) implies

$$\omega_\lambda^{(k)} \|u_\lambda\|_2^2 - \frac{\|u_\lambda\|_q^2}{4\theta_\lambda^{(k)}} \le \left(\omega_\lambda^{(k)} - \frac{1}{4\kappa_{q'}\theta_\lambda^{(k)}} \right) \|u_\lambda\|_2^2 \le \left(\omega_\lambda^{(k-1)} - \frac{1}{4\kappa_q \theta_\lambda^{(k-1)}} \right) \|u_\lambda\|_2^2$$
$$\le \omega_\lambda^{(k-1)} \|u_\lambda\|_2^2 - \frac{\|u_\lambda\|_q^2}{4\theta_\lambda^{(k)}}.$$

Thus, $F_{(k)}(u) \leq F_{(k-1)}(u)$ for all $u \in X$ and $k \geq k_0$. In particular, $F_{(k)} \geq K_\rho$ and, hence, coerciveness of K_ρ implies that $F_{(k)}$ is coercive as well. Thus, $F_{(k)}$ has a minimizer. Moreover, by (5.2) $J_{(k)}$ is strictly convex, and therefore the minimizer is unique. Invoking Theorem 5.2 yields the statement. ■

REMARK: Let us give explicit examples of sequences $\theta_\lambda^{(k)}$ and $\omega_\lambda^{(k)}$ satisfying the condition in Theorem 5.1. For $q \in \{2, \infty\}$ one may choose $\theta_\lambda^{(k)}$ increasing with k and satisfying (5.1), for instance $\theta_\lambda^{(k)} = k$. Then with $C > 1$ one chooses $\omega_\lambda^{(k)} = \frac{C}{4\theta_\lambda^{(k)}}$ and it is not difficult to verify (5.2) and (5.3).

For $q = 1$ one may choose $C > M$ and a sequence $\theta_\lambda^{(k)}$ such that $\theta_\lambda^{(k)} \geq \frac{(C-1)M}{CM-1}\theta_\lambda^{(k-1)}$ and (5.1) is satisfied, for instance

$$\theta_\lambda^{(k)} = \left(\frac{(C-1)M}{CM-1}\right)^k.$$

Then as before set $\omega_\lambda^{(k)} = \frac{C}{4\theta_\lambda^{(k)}}$ and again it is easy to verify (5.2) and (5.3).

5.2 Regularization results

By using the tool of Γ-convergence provided by Theorem 5.2 we can easily show two regularization results associated to the functional J. Let us first consider the functional

$$\mathcal{J}_\tau^1(u,v) := \mathcal{J}_{\theta,\rho,\omega;\tau}^1(u,v) := \|Tu - g|\mathcal{H}\|^2 + \tau\Phi_{\theta,\rho,\omega}^{(q)}(u,v).$$

We are interested in studying the behavior of the minimizers (u_τ^*, v_τ^*) of $\mathcal{J}_\tau^1$ for $\tau \to 0$.

We have the following straightforward result.

Lemma 5.3. *Let $q \in \{1, 2, \infty\}$ and assume that $\|T\| < 1$ and*

$$\inf_{\lambda\in\Lambda} 4\theta_\lambda(s_{\min} + \omega_\lambda) > \kappa_q \tag{5.6}$$

with $s_{\min} = \min \mathrm{Sp}(T^*T)$. For $\tau > 0$ let (u_τ^*, v_τ^*) be a minimizer of $\mathcal{J}_\tau^1$. Then $v_\tau^* = V_{\theta,\rho}^{(q)}(u_\tau^*)$.

Proof. Observe that both $\mathcal{J}_\tau^1$ and $J_{\frac{\theta}{\tau},\tau\rho,\tau\omega}^{(q)}$ are strictly convex and $\mathcal{J}_\tau^1(u,v) = J_{\frac{\theta}{\tau},\tau\rho,\tau\omega}^{(q)}(u,\tau v)$. Therefore (u_τ^*, v_τ^*) is the minimizer of $\mathcal{J}_\tau^1$ if and only if $(u_\tau^*, \tau v_\tau^*)$ is the minimizer of $J_{\frac{\theta}{\tau},\tau\rho,\tau\omega}^{(q)}$. Hence, $\tau v_\tau^* = V_{\frac{\theta}{\tau},\tau\rho}^{(q)}(u_\tau^*)$, due to the minimality of τv_τ^* for $J_{\frac{\theta}{\tau},\tau\rho,\tau\omega}^{(q)}(u_\tau^*, \cdot)$, and by definition, $V_{\frac{\theta}{\tau},\tau\rho}^{(q)}(u_\tau^*) = \tau V_{\theta,\rho}^{(q)}(u_\tau^*)$. We conclude that $v_\tau^* = V_{\theta,\rho}^{(q)}(u_\tau^*)$. ■

Proposition 5.4. *Let $q \in \{1, 2, \infty\}$ and assume that $\|T\| < 1$ and*

$$\inf_{\lambda\in\Lambda} 4\theta_\lambda\omega_\lambda > \kappa_q. \tag{5.7}$$

Furthermore suppose $\omega_\lambda \geq \gamma > 0$ for all $\lambda \in \Lambda$. Choose $u^\circ \in \ell_2(\Lambda, \mathbb{R}^M)$ and set $g = Tu^\circ$. Then for all $\tau > 0$ the functional $\mathcal{J}_{\theta,\omega,\rho;\tau}^1$ has a unique minimizer $(u_\tau^*, V_{\theta,\rho}^{(q)}(u_\tau^*))$. Choose a *sequence* $(\tau_n)_{n\in\mathbb{N}}$ *of positive reals which converges monotonically to* 0*. Then* $u^*_{\tau_n}$ *converges weakly to the unique solution* u^* *of the minimization problem*

$$\min_u \Phi^{(q)}_{\theta,\rho,\omega}(u, V^{(q)}_{\theta,\rho}(u)) \quad \text{subject to} \quad Tu = g. \tag{5.8}$$

Proof. We first argue that (5.8) has a unique solution. Existence follows from coerciveness, lower semicontinuity of $\Phi^{(q)}_{\theta,\omega,\rho}(u, V^{(q)}_{\theta,\rho}(u))$ and because $g = Tu^\circ$. Furthermore, the problem

$$\min_{(u,v)} \Phi^{(q)}_{\theta,\omega,\rho}(u, v) \quad \text{subject to} \quad Tu = g$$

has a unique solution $(u^\dagger, v^\dagger)$ since (5.7) implies strict convexity of $\Phi^{(q)}_{\theta,\omega,\rho}$. Then $v^\dagger = V^{(q)}_{\theta,\rho}(u^\dagger)$, and hence, $u^\dagger$ is also the unique minimizer of (5.8).

By the condition $\omega_\lambda \geq \gamma > 0$ for all $\lambda \in \Lambda$ we have $\mathcal{J}^1_\tau(u, v) \geq \tau\gamma\|u\|_2^2$, and hence $u \mapsto \mathcal{J}^1_\tau(u, V^{(q)}_{\theta,\rho}(u))$ is coercive with respect to the weak topology in ℓ_2. Hence, it has a minimizer and $(u, v) \mapsto \mathcal{J}^1_\tau(u, v)$ has a minimizer as well. Condition (5.7) implies strict convexity of $\mathcal{J}^1_\tau$, and thus the minimizer is unique. Now, we can estimate

$$\begin{aligned}\|u^*_{\tau_n}\|_2^2 &\leq \frac{1}{\gamma}\Phi^{(q)}_{\theta,\rho,\omega}(u^*_{\tau_n}, V^{(q)}_{\theta,\rho}(u^*_{\tau_n})) \leq \frac{1}{\gamma\tau_n}\mathcal{J}^1_{\tau_n}(u^*_{\tau_n}, V^{(q)}_{\theta,\rho}(u^*_{\tau_n})) \leq \frac{1}{\gamma\tau_n}\mathcal{J}^1_{\tau_n}(u^\dagger, V^{(q)}_{\theta,\rho}(u^\dagger)) \\ &= \frac{1}{\gamma}\Phi^{(q)}_{\theta,\rho,\omega}(u^\dagger, V^{(q)}_{\theta,\rho}(u^\dagger)).\end{aligned}$$

Therefore, the sequence $(u^*_{\tau_n})_{n\in\mathbb{N}}$ is uniformly bounded in $\ell_2(\Lambda, \mathbb{R}^M)$. Hence, there exists a subsequence $(u^*_{\tau_{n_j}})_{j\in\mathbb{N}}$ which converges weakly to $u^* \in \ell_2(\Lambda, \mathbb{R}^M)$. Let us denote again this subsequence by $(u^*_{\tau_n})_{n\in\mathbb{N}}$. Since it is bounded we can restrict our attention to the space $X = \{u \in \ell_2, \Phi^{(q)}_{\theta,\rho,\omega}(u, V^{(q)}_{\theta,\rho}(u)) \leq \Phi^{(q)}_{\theta,\rho,\omega}(u^\dagger, V^{(q)}_{\theta,\rho}(u^\dagger))\}$. With the same argument as above X is contained in the ℓ_2 ball of radius $1/\gamma$, and hence, X is metrizable when endowed with the weak topology induced by ℓ_2, see [10, Proposition 8.7].

Clearly, the term $\tau_n\Phi^{(q)}(u, V^{(q)}_{\theta,\rho}(u))$ converges monotonically to 0 for all $u \in X$. Hence, the sequence $(\mathcal{J}^1_{\tau_n}(u, \tau_n V^{(q)}_{\theta,\rho}(u)))_{n\in\mathbb{N}}$ converges pointwise and monotonically decreasing to $\|Tu - g|\mathcal{H}\|^2$. Since $\|Tu - g|\mathcal{H}\|^2$ is lower semicontinuous and trivially coercive in X (by weak compactness of X itself), we conclude by Theorem 5.2 that u^* minimizes $\|Tu - g|\mathcal{H}\|^2$, i.e., $Tu^* = g$.

Certainly

$$\Phi^{(q)}_{\theta,\rho,\omega}(u^*, V^{(q)}_{\theta,\rho}(u^*)) \geq \Phi^{(q)}_{\theta,\rho,\omega}(u^\dagger, V^{(q)}_{\theta,\rho}(u^\dagger)). \tag{5.9}$$

Since $\lim_{n\to\infty} V^{(q)}_{\theta,\rho}(u^*_{\tau_n})_\lambda = V^{(q)}_{\theta,\rho}(u^*)_\lambda$, by Fatou's lemma, we can estimate

$$\Phi^{(q)}_{\theta,\rho,\omega}(u^*, V^{(q)}_{\theta,\rho}(u^*)) \leq \liminf_{n\to\infty} \Phi^{(q)}_{\theta,\rho,\omega}(u^*_{\tau_n}, V^{(q)}_{\theta,\rho}(u^*_{\tau_n})).$$

Since $\Phi^{(q)}_{\theta,\rho,\omega}(u^*_{\tau_n}, V^{(q)}_{\theta,\rho}(u^*_{\tau_n}))$ is bounded, we may again pass to a subsequence of τ_n (labelled again τ_n) so that the right hand side above converges, and

$$\Phi^{(q)}_{\theta,\rho,\omega}(u^*, V^{(q)}_{\theta,\rho}(u^*)) \leq \lim_{n\to\infty} \Phi^{(q)}_{\theta,\rho,\omega}(u^*_{\tau_n}, V^{(q)}_{\theta,\rho}(u^*_{\tau_n})).$$

We have shown before (actually by definition of X) that

$$\Phi^{(q)}_{\theta,\rho,\omega}(u^*_{\tau_n}, V^{(q)}_{\theta,\rho}(u^*_{\tau_n})) \le \frac{1}{\tau_n}\mathcal{J}^1_{\tau_n}(u^*_{\tau_n}, V^{(q)}_{\theta,\rho}(u^*_{\tau_n})) \le \Phi^{(q)}_{\theta,\rho,\omega}(u^\dagger, V^{(q)}_{\theta,\rho}(u^\dagger))$$

and therefore $\Phi^{(q)}_{\theta,\rho,\omega}(u^*, V^{(q)}_{\theta,\rho}(u^*)) \le \Phi^{(q)}_{\theta,\rho,\omega}(u^\dagger, V^{(q)}_{\theta,\rho}(u^\dagger))$. Combining the last inequality with (5.9) we obtain that u^* solves the minimization problem (5.8) and thus coincide with $u^\dagger$. Hence, any subsequence of $u^*_{\tau_n}$ possesses another subsequence which converges to $u^\dagger$, and we conclude that $u^*_{\tau_n}$ itself must converge to $u^\dagger$. ∎

A similar result can be achieved as well by considering a slightly different sequence of functionals. Define

$$\Psi^{(q)}_{\omega}(u,v) := \sum_{\lambda\in\Lambda} v_\lambda \|u_\lambda\|_q + \sum_{\lambda\in\Lambda} \omega_\lambda \|u_\lambda\|_2^2$$

and

$$\mathcal{J}^2_\tau(u,v) \;:=\; \mathcal{J}^2_{\theta,\omega,\rho;\tau}(u,v) \;:=\; \|Tu - g|\mathcal{H}\|^2 + \tau\Psi^{(q)}_{\omega}(u,v) + \sum_{\lambda} \theta_\lambda(\rho_\lambda - v_\lambda)^2. \tag{5.10}$$

Similarly to the arguments above we can show the following results.

Lemma 5.5. *Let $q \in \{1, 2, \infty\}$ and assume that $\|T\| < 1$ and*

$$\inf_{\lambda\in\Lambda} 4\theta_\lambda(s_{\min} + \omega_\lambda) > \kappa_q \tag{5.11}$$

with $s_{\min} = \min \mathrm{Sp}(T^*T)$. For $\tau > 0$ let (u^*_τ, v^*_τ) be a minimizer of $\mathcal{J}^2_\tau$. Then $v^*_\tau = V^{(q)}_{\frac{\theta}{\tau},\rho}(u^*_\tau)$.

Proposition 5.6. *Let $q \in \{1, 2, \infty\}$ and assume that $\|T\| < 1$ and*

$$\inf_{\lambda\in\Lambda} 4\theta_\lambda(s_{\min} + \omega_\lambda) > \kappa_q \tag{5.12}$$

with $s_{\min} = \min \mathrm{Sp}(T^*T)$. We furthermore assume that $\omega_\lambda \ge \gamma > 0$ for all $\lambda \in \Lambda$. Choose $u^\circ \in \ell_2(\Lambda, \mathbb{R}^M)$ and set $g = Tu^\circ$. Then for all $\tau > 0$ the functional $\mathcal{J}^2_\tau = \mathcal{J}^2_{\theta,\omega,\rho;\tau}$ has a unique minimizer $(u^*_\tau, V^{(q)}_{\frac{\theta}{\tau},\rho}(u^*_\tau))$. Let $(\tau_n)_{n\in\mathbb{N}}$ be a sequence of positive reals which converges monotonically to 0. Then $u^*_{\tau_n}$ converges weakly, and its limit u^* uniquely solves the minimization problem

$$\min_u \Psi^{(q)}_{\omega}(u,\rho) \quad \text{subject to} \quad Tu = g. \tag{5.13}$$

Proof. Exactly as in the proof of Proposition 5.4 one shows uniqueness of the minimizer u^*_τ. Let $u^\dagger$ be a solution of the minimization problem (5.13), which is unique since $\omega_\lambda \ge \gamma > 0$ implies that $u \mapsto \Psi^{(q)}_{\omega}(u,\rho)$ is strictly convex. We can estimate

$$\begin{aligned}\|u^*_{\tau_n}\|_2^2 &\le \frac{1}{\gamma}\Psi^{(q)}_{\omega}(u^*_{\tau_n}, V^{(q)}_{\frac{\theta}{\tau_n},\rho}(u^*_{\tau_n})) \le \frac{1}{\gamma\tau_n}\mathcal{J}^2_{\tau_n}(u^*_{\tau_n}, V^{(q)}_{\frac{\theta}{\tau_n},\rho}(u^*_{\tau_n})) \le \frac{1}{\gamma\tau_n}\mathcal{J}^2_{\tau_n}(u^\dagger,\rho)) \\ &= \frac{1}{\gamma}\Psi^{(q)}_{\omega}(u^\dagger,\rho).\end{aligned}$$

Therefore, the sequence $(u^*_{\tau_n})_{n\in\mathbb{N}}$ is uniformly bounded in $\ell_2(\Lambda,\mathbb{R}^M)$ and we can extract a subsequence $(u^*_{\tau_{n_j}})_{j\in\mathbb{N}}$ which converges weakly to $u^*\in\ell_2(\Lambda,\mathbb{R}^M)$. Let us denote again such subsequence by $(u^*_{\tau_n})_{n\in\mathbb{N}}$. We may restrict our attention to the space

$$X := \left\{u\in\ell_2, \|u\|_2^2 \le \gamma^{-1}\Psi^{(q)}_\omega(u^\dagger,\rho)\right\} \cap \left\{u\in\ell_2, \exists v \text{ such that } \mathcal{J}^2_1(u,v)<\infty\right\}.$$

By definition X is contained in an ℓ_2-ball and hence, metrizable when endowed with the weak topology induced from ℓ_2. It is easy to see that $\mathcal{J}^2_\tau(u,v)<\infty$ for arbitrary $\tau>0$ if $\mathcal{J}^2_1(u,v)<\infty$. Moreover, for fixed $u\in X$ the weight $v'=V_{\frac{\theta}{\tau},\rho}(u)$ minimizes $v\mapsto\mathcal{J}^2_\tau(u,v)$, i.e., $\mathcal{J}^2_\tau(u,V_{\frac{\theta}{\tau},\rho}(u))\le\mathcal{J}^2_\tau(u,v)$. In particular, $u^*_{\tau_n}$ is contained in X for all n.

Now, a straightforward computation shows that

$$\tau_n^2\sum_{\lambda\in\Lambda}\frac{\|u_\lambda\|_q^2}{4\theta_\lambda} = \sum_{\lambda\in\Lambda}\theta_\lambda(\rho_\lambda - V^{(q)}_{\frac{\theta}{\tau_n},\rho}(u)_\lambda)^2 \le \mathcal{J}^1_{\tau_n}(u,V^{(q)}_{\frac{\theta}{\tau_n},\rho}(u))<\infty$$

for all $u\in X$ and n large enough, hence the first term converges monotonically to 0 as $n\to\infty$. It follows that the sequence $(\mathcal{J}^2_{\tau_n}(u,V^{(q)}_{\frac{\theta}{\tau_n},\rho}(u)))_{n\in\mathbb{N}}$ monotonically converges to $\|Tu-g|\mathcal{H}\|^2$ for all $u\in X$.

Since $\|Tu-g|\mathcal{H}\|^2$ is lower semicontinuous and coercive in X it follows from Theorem 5.2 that u^* minimizes $\|Tu-g|\mathcal{H}\|^2$, i.e., $Tu^*=g$.

Clearly,

$$\Psi^{(q)}_\omega(u^*,\rho)\ge\Psi^{(q)}_\omega(u^\dagger,\rho). \tag{5.14}$$

Using $\lim_n V^{(q)}_{\frac{\theta}{\tau_n},\rho}(u^*_{\tau_n})_\lambda=\rho_\lambda$ we can show (by possibly passing to a subsequence of (τ_n))

$$\Psi^{(q)}_\omega(u^*,\rho) \;=\; \Phi^{(q)}_{\theta,\rho,\omega}(u^*,\rho) \;\le\; \lim_{n\to\infty}\Phi^{(q)}_{\theta,\rho,\omega}(u^*_{\tau_n},V^{(q)}_{\frac{\theta}{\tau_n},\rho}(u^*_{\tau_n})).$$

We have shown before that $\Phi^{(q)}_{\theta,\rho,\omega}(u^*_{\tau_n},V^{(q)}_{\frac{\theta}{\tau_n},\rho}(u^*_{\tau_n})) \le \frac{1}{\tau_n}\mathcal{J}^2_{\tau_n}(u^*_{\tau_n},V^{(q)}_{\frac{\theta}{\tau_n},\rho}(u^*_{\tau_n})) \le \Psi^{(q)}_\omega(u^\dagger,\rho)$ and therefore we have $\Psi^{(q)}_\omega(u^*,\rho)\le\Psi^{(q)}_\omega(u^\dagger,\rho)$. Together with (5.14) we obtain $u^*=u^\dagger$, and with the same argument as in the proof of Proposition 5.4 we argue that the full sequence u_{τ_n} converges weakly to $u^\dagger$. ■

A Appendix

Proof of Theorem 3.2

The proof uses subdifferentials. This requires to formally extend the function $G_{\theta,\rho,\omega;z}$ to $\mathbb{R}^M\times\mathbb{R}$ by setting $G_{\theta,\rho,\omega;z}(x,y)=\infty$ if $y<0$. In [24] the following characterization was provided.

Lemma A.1. *Let $(u,v)\in\mathbb{R}^M\times\mathbb{R}_+$. Then $(\xi,\eta)\in\mathbb{R}^M\times\mathbb{R}$ is contained in the subdifferential $\partial G^{(q)}_{\theta,\rho,\omega;z}(x,y)$ if and only if*

$$\begin{aligned}\xi &\in 2(1+\omega)u-2z+v\partial\|\cdot\|_q(u),\\ \eta &\in \|u\|_q\partial s^+(v)+2\theta(v-\rho),\end{aligned}$$

where $s^+(v) := v$ for $v \geq 0$ and $s^+(v) = \infty$ for $v < 0$.

REMARK: We recall that the subdifferential of the q-norm on $\mathbb{R}^M$ is given as follows.

- If $1 < q < \infty$ then

$$\partial \| \cdot \|_q(x) \;=\; \begin{cases} B^{q'}(1) & \text{if } x = 0, \\ \left\{ \left(\frac{|x^\ell|^{q-1} \operatorname{sign}(x^\ell)}{\|x\|_q^{q-1}} \right)_{\ell=1}^M \right\} & \text{otherwise,} \end{cases}$$

  where $B^{q'}(1)$ denotes the ball of radius 1 in the dual norm, i.e., in $\ell_{q'}$ with $\frac{1}{q} + \frac{1}{q'} = 1$.

- If $q = 1$ then

$$\partial \| \cdot \|_1(x) \;=\; \{\xi \in \mathbb{R}^M : \; \xi_\ell \in \partial | \cdot |(x^\ell), \ell = 1, \dots, M\} \tag{A.1}$$

  where $\partial | \cdot |(z) = \{\operatorname{sign}(z)\}$ if $z \neq 0$ and $\partial | \cdot |(0) = [-1, 1]$.

- If $q = \infty$ then

$$\partial \| \cdot \|_\infty(x) \;=\; \begin{cases} B^1(1) & \text{if } x = 0, \\ \operatorname{conv}\{(\operatorname{sign}(x^\ell) e_\ell : |x^\ell| = \|x\|_\infty\} & \text{otherwise,} \end{cases} \tag{A.2}$$

  where conv A denotes the convex hull of a set A and e_ℓ the ℓ-th canonical unit vector in $\mathbb{R}^M$.

Proof of Theorem 3.2. First observe that $G_{\theta,\rho,\omega;z}$ is strictly convex, continuous on its domain $\mathbb{R}^M \times \mathbb{R}_+$, and bounded from below, further $G_{\theta,\rho,\omega;z}(x, y) \to \infty$ when $\|x\|_2 + |y| \to \infty$. Thus, there exists a unique minimizer. Hence, we have to prove that $0 \in \partial G_{\theta,\rho,\omega;z}(u, v)$. It is straightforward to see that once $\|u\|_q$ is known then v is given by (3.9).

From here on we have to distinguish between the different q. Let us start with the easiest case $q = 2$. Assume $u \neq 0$ and $\|u\|_2 \leq 2\theta\rho$. By the characterization of the subdifferential in Lemma A.1 it follows that $v = \rho - \frac{\|u\|_2}{2\theta}$, and $0 \in 2(1+\omega)u - 2z + (\rho - \frac{\|u\|_2}{2\theta})\partial \| \cdot \|_2(u)$. Since $u \neq 0$, we have $\partial \| \cdot \|_2(u) = \{\frac{u}{\|u\|_2}\}$ and $0 = 2(1+\omega)u - 2z + (\frac{\rho}{\|u\|_2} - \frac{1}{2\theta})u$. A straightforward computation gives

$$z \;=\; \left((1+\omega) + \frac{\rho}{2\|u\|_2} - \frac{1}{4\theta} \right) u$$

and hence

$$\|z\|_2 \;\;=\;\; \left((1+\omega) + \frac{\rho}{2\|u\|_2} - \frac{1}{4\theta} \right) \|u\|_2 = \left((1+\omega) - \frac{1}{4\theta} \right) \|u\|_2 + \frac{\rho}{2}.$$

Since by assumption $4\theta(1+\omega) > 1$ we find that

$$\|u\|_2 = \frac{\|z\|_2 - \rho/2}{(1+\omega) - \frac{1}{4\theta}}.$$

The latter equivalence makes sense only if $\|z\|_2 - \rho/2 > 0$, otherwise we would have a contradiction to $u \neq 0$.

If $u = 0$ then $v = \rho$ and necessarily $\|z\|_2 \le \rho/2$. This proves that $u = 0$ if and only if $\|z\|_2 \le \rho/2$. So let us assume then $\|z\|_2 - \rho/2 > 0$. By the computations done above we obtain

$$z = \left((1+\omega) + \frac{\rho}{2\frac{\|z\|_2 - \rho/2}{(1+\omega) - \frac{1}{4\theta}}} - \frac{1}{4\theta} \right) u,$$

which is equivalent to

$$u = \frac{\|z\|_2 - \rho/2}{(1+\omega - \frac{1}{4\theta})\|z\|_2} z = (1+\omega)^{-1} \frac{4\theta(1+\omega)}{4\theta(1+\omega) - 1} \frac{\|z\|_2 - \rho/2}{\|z\|_2} z.$$

Due to the assumption $\|u\|_2 \le 2\theta\rho$, this relation can only hold if $\|z\|_2 \le 2\theta(1+\omega)\rho$.

Let us finally assume that $\|u\|_2 > 2\theta\rho$. Then $v = 0$ and it is straightforward to check that $u = (1+\omega)^{-1} z$, and $\|z\|_2 \ge 2\theta(1+\omega)\rho$. Summarizing the results, and considering the definition of $h^{(2)}_{\theta(1+\omega),\rho}$ we have

$$u = (1+\omega)^{-1} h^{(2)}_{\theta(1+\omega),\rho}(z)$$

as claimed.

Let us turn to the case $q = 1$. We assume first $u \ne 0$ and $\|u\|_1 \le 2\theta\rho$. By Lemma A.1 it follows that $v = \rho - \frac{\|u\|_1}{2\theta}$, and $0 \in 2(1+\omega)u - 2z + (\rho - \frac{\|u\|_1}{2\theta})\partial\|\cdot\|_1(u)$. The latter condition implies

$$u_\ell = (1+\omega_\lambda)^{-1} \begin{cases} 0, & |z_\ell| \le \rho/2 - \frac{\|u\|_1}{4\theta}, \\ z_\ell - \operatorname{sign}(z_\ell)\left(\rho/2 - \frac{\|u\|_1}{4\theta}\right), & |z_\ell| > \rho/2 - \frac{\|u\|_1}{4\theta}. \end{cases} \tag{A.3}$$

Thus, we need to determine $\|u\|_1$. Let $\ell \in \{1, \dots, M\}$ and assume $u^\ell \ne 0$. Then we have $0 = 2(1+\omega)u - 2z + (\rho - \frac{\|u\|_1}{2\theta}) \operatorname{sign}(u^\ell)$, hence $z^\ell = (1+\omega)u^\ell + \left(\frac{\rho}{2} - \frac{\|u\|_1}{4\theta}\right) \operatorname{sign}(u^\ell)$ and $|z^\ell| = (1+\omega)|u_\ell| + \left(\frac{\rho}{2} - \frac{\|u\|_1}{4\theta}\right)$. Denoting $S = \operatorname{supp}(u) = \{\ell : u_\ell \ne 0\}$ and $n = \#S$ we obtain

$$\sum_{\ell \in S} |z_\ell| = (1+\omega)\|u\|_1 + n\left(\frac{\rho}{2} - \frac{\|u\|_1}{4\theta}\right), \tag{A.4}$$

Thus, we need to determine S and n in order to compute $\|u\|_1$, i.e.,

$$\|u\|_1 = \frac{4\theta}{4\theta(1+\omega) - n} \left(\sum_{\ell \in S} |z_\ell| - \frac{n\rho}{2} \right) =: v_S(z). \tag{A.5}$$

Summarizing the conditions needed so far, the set S (of cardinality n) has to satisfy

$$|z_\ell| > \rho/2 - \frac{\sum_{\ell \in S} |z_\ell| - \frac{n\rho}{2}}{4\theta(1+\omega) - n}, \quad \text{for all } \ell \in S, \tag{A.6}$$

$$|z_\ell| \le \rho/2 - \frac{\sum_{\ell \in S} |z_\ell| - \frac{n\rho}{2}}{4\theta(1+\omega) - n}, \quad \text{for all } \ell \notin S, \tag{A.7}$$

and

$$0 \le v_S(z) = \|u\|_1 \le 2\theta\rho \tag{A.8}$$

by the initial assumption $\|u\|_1 \leq 2\theta\rho$. By (A.6) and (A.7), S has to contain the n largest absolute value coefficients of z. Thus, if the entries of z are ordered such that $|z_{\ell_1}| \geq |z_{\ell_2}| \geq \ldots \geq |z_{\ell_M}|$ then it suffices to find n such that

$$\frac{1}{n}\sum_{j=1}^{n} |z_{\ell_j}| \;\geq\; \frac{\rho}{2}, \tag{A.9}$$

$$\sum_{j=1}^{n} |z_{\ell_j}| \;\leq\; 2\theta(1+\omega)\rho, \tag{A.10}$$

and

$$|z_{\ell_n}| \;>\; \frac{\rho}{2} - \frac{\sum_{j=1}^{n} |z_{\ell_j}| - \frac{n\rho}{2}}{4\theta(1+\omega) - n}, \tag{A.11}$$

$$|z_{\ell_{n+1}}| \;\leq\; \frac{\rho}{2} - \frac{\sum_{j=1}^{n} |z_{\ell_j}| - \frac{n\rho}{2}}{4\theta(1+\omega) - n}, \tag{A.12}$$

where the last condition is void if $n = M$. Note that condition (A.10) is a straightforward consequence of (A.5) and $\|u\|_1 \leq 2\theta\rho$.

Observe that the sequence $n \mapsto n^{-1}\sum_{j=1}^{n} |z_{\ell_{n_j}}|$ is decreasing with n by the ordering of $|z_{\ell_j}|$. Thus, if $\rho/2 > |z_{\ell_1}| = \|z\|_\infty$ then condition (A.9) cannot be satisfied for any $n \in \{1,\ldots,M\}$. In this case the initial assumption was consequently wrong, and hence, either $u = 0$ or $\|u\|_1 > 2\theta\rho$. If $\|u\|_1 > 2\theta\rho$ then $v = 0$, and hence, $u = (1+\omega)^{-1}(z)$, i.e., $\|z\|_1 = (1+\omega)\|u\|_1 \geq 2\theta(1+\omega)\rho$ which contradicts $\|z\|_\infty < \rho/2$ as $\|z\|_1 \leq M\|z\|_\infty < M\rho/2 < 2\theta(1+\omega)\rho$ by the assumption $\theta(1+\omega) > M/4$. Thus, we conclude that $u = 0$ if $\|z\|_\infty < \rho/2$.

Now assume that $\|z\|_1 > 2\theta(1+\omega)\rho$. First note that then $u = 0$ is not possible. Indeed, if $u = 0$ then $v = \rho$ and hence, $z \in \rho/2\partial\|\cdot\|_1(0) = B^\infty(\rho/2)$. Hence, $\|z\|_\infty \leq \rho/2$ which contradicts $\|z\|_1 > 2\theta(1+\omega)$ by the same reasoning as above. We now argue that also $\|u\|_1 \leq 2\theta\rho$ is not possible. Clearly, if $\|z\|_1 > 2\theta(1+\omega)\rho$ then (A.10) is not satisfied for $n = M$. However, there might exist $n = m < M$ for which (A.10) is satisfied. In this case it suffices to show that condition (A.12) is never satisfied for $n = 1,\ldots,m$. Indeed, for $n \leq m$ we estimate the right hand side of (A.12) as

$$\begin{aligned}
\frac{\rho}{2} - \frac{\sum_{j=1}^{n} |z_{\ell_j}| - \frac{n\rho}{2}}{4\theta(1+\omega) - n} &= \frac{\rho}{2} - \frac{\|z\|_1 - \sum_{j=n+1}^{M} |z_{\ell_j}| - \frac{n\rho}{2}}{4\theta(1+\omega) - n} \\
&< \frac{\rho}{2} - \frac{2\theta(1+\omega)\rho - (M-n)|z_{\ell_{n+1}}| - \frac{n\rho}{2}}{4\theta(1+\omega) - n} = \frac{M-n}{4\theta(1+\omega) - n}|z_{\ell_{n+1}}| < |z_{\ell_{n+1}}|.
\end{aligned}$$

Here we used the ordering of the $|z_{\ell_j}|$ and $4\theta(1+\omega) > M$. Thus, (A.12) cannot be satisfied and, hence, we necessarily have $\|u\|_1 > 2\theta\rho$. As already mentioned above we obtain $u = (1+\omega)^{-1}z$ in this case.

It remains to treat the case $\|z\|_\infty > \rho/2$ and $\|z\|_1 \leq 2\theta(1+\omega)$. In this case it it is not possible that $u = 0$ since then $v = \rho$ and, hence, $z \in B^\infty(\rho/2)$, i.e., $\|z\|_\infty \leq \rho/2$, as already noted above. Also $\|u\|_1 > 2\theta\rho$ cannot hold since this would imply $u = (1+\omega)^{-1}z$ and consequently $\|z\|_1 = (1+\omega)\|u\|_1 > 2\theta\rho(1+\omega)$. This means that we are in the situation assumed in the beginning of the proof for $q = 1$. Since u exists and and is unique also its support is unique and there must exist a unique n satisfying (A.9), (A.10), (A.11) and (A.12). Once n is known, the support S of u corresponds to the indices of the n largest entries of z and $\|u\|_1$ is given by (A.5), while the entries of u_ℓ are determined by (A.3). Considering the definition of $t_n(z)$ in (3.5) (with θ replaced by $\theta(1+\omega)$) we deduce that

$$u = (1+\omega)^{-1} h^{(1)}_{\theta(1+\omega),\rho}(z)$$

for all the cases as claimed.

Let us finally consider $q = \infty$. Let us assume for the moment that $u \neq 0$ and $\|u\|_\infty \leq 2\theta\rho$. Then $v = \rho - \frac{\|u\|_\infty}{2\theta}$. Let S be the set of indices ℓ for which $|u_\ell| = \|u\|_\infty$. We enumerate them by $\ell_1, \ldots, \ell_n$. For simplicity we further assume that entries $z_{\ell_1}, \ldots, z_{\ell_n}$ are positive (the other cases can be treated similarly by taking into account the corresponding signs). Then the numbers $u_{\ell_1}, \ldots, u_{\ell_n}$ are also positive since choosing them with opposite signs would increase the function $G_{\theta,\rho,\omega;z}$. From Lemma A.1 and the characterization of $\partial\|\cdot\|_\infty(u)$ we see that $2(u_\ell(1+\omega) - z_\ell) = 0$ for the u_ℓ not giving the maximum, i.e.,

$$u_\ell = (1+\omega)^{-1} z_\ell \quad \text{for } \ell \notin S.$$

If $n := \#S = 1$, i.e., the maximum is attained at only one entry, then for the corresponding $\ell \in S$ we obtain by Lemma A.1, $0 = 2(1+\omega)u_\ell - 2z_\ell + \rho - \frac{\|u\|_\infty}{2\theta}$, i.e.,

$$u_\ell = (1+\omega)^{-1}\left(z_\ell - \left(\frac{\rho}{2} - \frac{\|u\|_\infty}{4\theta}\right)\right).$$

As $u_\ell = \|u\|_\infty$ this necessarily implies $z_\ell > z_{\ell'}$ for $\ell' \notin S = \{\ell\}$, i.e., $|z_\ell| = \|z\|_\infty$. Moreover, solving for u_ℓ yields

$$u_\ell \;=\; \frac{4\theta}{4\theta(1+\omega)-1}\left(z_\ell - \rho/2\right).$$

Since $u_\ell > 0$ and $u_\ell \leq 2\theta\rho$ this necessarily requires $z_\ell = \|z\|_\infty > \rho/2$ and $\|z\|_\infty \leq 4\theta(1+\omega)\rho$. The realization of the maximum only at u_ℓ is valid only if $u_{\ell'} < u_\ell$ for all $\ell' \notin S = \{\ell\}$, i.e.,

$$z_{\ell'} < \frac{4\theta(1+\omega)}{4\theta(1+\omega)-1}\left(\|z\|_\infty - \rho/2\right).$$

Otherwise we may assume that $n = \#S > 1$ and we put

$$t := \|u\|_\infty = u_\ell \quad \text{for all } \ell \in S.$$

By the characterization in Lemma A.1 and the explicit form of $\partial\|\cdot\|_\infty(u)$ we then have

$$\begin{aligned} 2t - 2z_{\ell_j} \;&=\; -\left(\rho - \frac{t}{2\theta}\right)a_j, \quad j = 1,\ldots,n-1,\\ 2t - 2z_{\ell_n} \;&=\; -\left(\rho - \frac{t}{2\theta}\right)\left(1 - \sum_{k=1}^{n-1} a_k\right) \end{aligned}$$

for some numbers $a_1, \ldots, a_{n-1} \in [0,1]$ satisfying $\sum_j a_j \leq 1$. This is a system of n nonlinear equations in t and $a_1, \ldots, a_{n-1}$. We proceed to its explicit solution by following two steps:

- We solve first the linear problem

$$\begin{aligned} 2(1+\omega)t - 2z_{\ell_j} &= -va_j, \quad j = 1, \ldots, n-1, \\ 2(1+\omega)t - 2z_{\ell_n} &= -v\left(1 - \sum_{k=1}^{n-1} a_k\right). \end{aligned}$$

- The solution $t = T(v, z_{\ell_1}, \ldots, z_{\ell_n})$ of the linear problem depends on the data $v, z_{\ell_1}, \ldots, z_{\ell_n}$. Since $v = \left(\rho - \frac{t}{2\theta}\right)$ we can find the solution of the nonlinear system by solving the fixed point equation

$$t = T(\rho - \frac{t}{2\theta}, z^{\ell_1}, \ldots, z^{\ell_n}).$$

So, let us solve the linear problem. To this end we follow the computations in [24, Lemma 4.2]. The linear system can be reformulated in matrix form as follows:

$$\underbrace{\begin{pmatrix} 1+\omega & v/2 & 0 & 0 & \cdots & 0 \\ 1+\omega & 0 & v/2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1+\omega & -v/2 & -v/2 & -v/2 & \cdots & -v/2 \end{pmatrix}}_{:=B} \begin{pmatrix} t \\ a_1 \\ \vdots \\ a_{n-1} \end{pmatrix} = \begin{pmatrix} z_{\ell_1} \\ \vdots \\ z_{\ell_{n-1}} \\ z_{\ell_n} - v/2 \end{pmatrix}.$$

Denoting the matrix on the left hand side by B, a simple computation verifies that

$$B^{-1} = \frac{1}{n} \begin{pmatrix} (1+\omega)^{-1} & (1+\omega)^{-1} & (1+\omega)^{-1} & \cdots & (1+\omega)^{-1} \\ \frac{2(n-1)}{v} & -\frac{2}{v} & -\frac{2}{v} & \cdots & -\frac{2}{v} \\ -\frac{2}{v} & \frac{2(n-1)}{v} & -\frac{2}{v} & \cdots & -\frac{2}{v} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{2}{v} & \cdots & -\frac{2}{v} & \frac{2(n-1)}{v} & -\frac{2}{v} \end{pmatrix}.$$

Then we can compute explicitly the solution t by

$$t = \frac{1}{n(1+\omega)} \left(\sum_{j=1}^{n} z_{\ell_j} - \frac{v}{2} \right).$$

By substituting $v = \rho - \frac{\|u\|_\infty}{2\theta} = \rho - \frac{t}{2\theta}$ into the last expression and solving the equation for t we obtain

$$u_{\ell_1} = \cdots = u_{\ell_n} = t = \frac{4\theta}{4\theta(1+\omega)n - 1} \left(\sum_{j=1}^{n} z_{\ell_j} - \frac{\rho}{2} \right).$$

Since $\|u\|_\infty = t$ and $0 < \|u\|_\infty \leq 2\theta\rho$ by the initial assumption this requires

$$\sum_{j=1}^{n} |z_{\ell_j}| > \rho/2 \tag{A.13}$$

and

$$\frac{1}{n}\sum_{j=1}^{n}|z_\ell| \leq 2\rho\theta(1+\omega). \tag{A.14}$$

The solution of the linear system gives also $a_j = \frac{2}{nv}\left(v/2 + (n-1)z_{\ell_j} - \sum_{k\in\{1,\dots,n\}\setminus\{j\}} z_{\ell_k}\right)$. We require $a_j \geq 0$ and $1 - \sum_{j=1}^{n-1} a_j \geq 0$. We have $a_j \geq 0$ if and only if

$$z_{\ell_j} \geq \frac{1}{n-1}\left(\sum_{k\in\{1,\dots,n\}\setminus\{j\}} z_{\ell_k} - v/2\right).$$

By substituting $v = \left(\rho - \frac{t}{2\theta}\right)$ and recalling the value of t as just computed above we obtain

$$z_{\ell_j} \geq \frac{4\theta(1+\omega)}{4\theta(1+\omega)(n-1)-1}\left(\sum_{k\in\{1,\dots,n\}\setminus\{j\}} z_{\ell_k} - \rho/2\right). \tag{A.15}$$

A direct computation also shows that $\sum_{j=1}^{n-1} a_j = \frac{n-1}{n} + \frac{2}{nv}\left(\sum_{j=1}^{n-1} z_{\ell_j} - (n-1)z_{\ell_n}\right)$. Thus, it holds $1 - \sum_{j=1}^{n-1} a_j \geq 0$ if and only if

$$z_{\ell_n} \geq \frac{1}{n-1}\left(\sum_{j=1}^{n-1} z_{\ell_j} - v/2\right).$$

Again the substitution of $v = \left(\rho - \frac{t}{2\theta}\right)$ gives

$$z_{\ell_n} \geq \frac{4\theta(1+\omega)}{4\theta(1+\omega)(n-1)-1}\left(\sum_{j=1}^{n-1} z_{\ell_j} - \rho/2\right). \tag{A.16}$$

The initial assumption that the maximum of n is attained precisely at $u_{\ell_1}, \dots, u_{\ell_n}$ can be true only if

$$z_{\ell'} = (1+\omega)u_{\ell'} < (1+\omega)t = \frac{4\theta(1+\omega)}{4\theta(1+\omega)n-1}\left(\sum_{j=1}^{n} z_{\ell_j} - \rho/2\right) \quad \text{for all } \ell' \notin S. \tag{A.17}$$

By combining this condition with (A.15) and (A.16) we deduce that S necessarily contains the indices ℓ_j corresponding to the largest coefficients of z. Thus, we may assume that the indices are ordered such that $|z_{\ell_1}| \geq |z_{\ell_2}| \geq \ldots \geq |z_{\ell_M}|$.

Summarizing what we have deduced so far, in particular, (A.13), (A.14), (A.16) and (A.17), the conditions $u \neq 0$ and $\|u\|_1 \leq 2\theta(1+\omega)\rho$ hold if and only if there exists $n \in \{1, \dots, M\}$ such

$$\sum_{j=1}^{n}|z_{\ell_j}| > \rho/2, \tag{A.18}$$

$$\frac{1}{n}\sum_{j=1}^{n}|z_{\ell_j}| \leq 2\rho\theta(1+\omega), \tag{A.19}$$

and

$$|z_{\ell_{n+1}}| \;<\; \frac{4\theta(1+\omega)}{4\theta(1+\omega)n-1}\left(\sum_{j=1}^{n}|z_{\ell_j}| - \frac{\rho}{2}\right) \;=\; s_n(z), \tag{A.20}$$

$$|z_{\ell_n}| \;\geq\; \frac{4\theta(1+\omega)}{4\theta(1+\omega)(n-1)-1}\left(\sum_{j=1}^{n-1}|z_{\ell_j}| - \frac{\rho}{2}\right) \;=\; s_{n-1}(z), \tag{A.21}$$

where the first condition is only considered if $n \leq M-1$ and the last condition if $n > 1$.

Now assume that $\|z\|_1 \leq \rho/2$. Then clearly, there exists no $n \in \{1,\dots,M\}$ such that (A.18) is satisfied. Thus, either $u = 0$ or $\|u\|_\infty > 2\theta\rho$. If $\|u\|_\infty > 2\theta\rho$ then $v = 0$ and $u = (1+\omega)^{-1}z$. Consequently, $\|z\|_\infty = (1+\omega)^{-1}\|u\|_\infty > 2\rho\theta(1+\omega)$ which yields a contradiction to the assumption as $\|z\|_\infty \leq \|z\|_1 \leq \rho/2 < 2\rho\theta(1+\omega)$ by (3.8). Thus, $u = 0$ if $\|z\|_1 \leq \rho/2$.

We assume next that $\|z\|_\infty > 2\theta\rho(1+\omega)$. In this case condition (A.19) is certainly not satisfied for $n = 1$. However, there might exist $n > 1$ such that $\sum_{j=1}^{n-1}|z_{\ell_j}| > 2\rho\theta(1+\omega)(n-1)$ but $\sum_{j=1}^{n}|z_{\ell_j}| \leq 2\rho\theta(1+\omega)n$. A straightforward computation shows then that $|z_{\ell_n}| < 2\theta\rho(1+\omega)$. Furthermore,

$$\begin{aligned} s_{n-1}(z) \;&=\; \frac{4\theta(1+\omega)}{4\theta(1+\omega)(n-1)-1}\left(\sum_{j=1}^{n-1}|z_{\ell_j}| - \frac{\rho}{2}\right) \\ &>\; \frac{4\theta(1+\omega)}{4\theta(1+\omega)(n-1)-1}(2\theta\rho(1+\omega)(n-1)-\rho/2) \;=\; 2\theta\rho(1+\omega). \end{aligned}$$

Hence, condition (A.21) is not satisfied for this particular n. We now argue that then also for $n' > n$ (A.21) cannot be satisfied. To this end we claim that $|z_{\ell_m}| \geq s_m(z)$ implies $s_m(z) \geq s_{m-1}(z)$ for arbitrary m. Then $|z_{\ell_{n+1}}| \geq s_n(z)$ would imply $|z_{\ell_n}| \geq |z_{\ell_{n+1}}| \geq s_n(z) \geq s_{n-1}(z)$, a contradiction to what we have just shown, and by induction (A.21) cannot hold for arbitrary $n' > n$. To prove the claim we estimate

$$\begin{aligned} &\frac{1}{4\theta(1+\omega)}(s_n(z) - s_{n-1}(z)) \\ &= \left(\frac{1}{4\theta(1+\omega)n-1} - \frac{1}{4\theta(1+\omega)(n-1)-1}\right)\left(\sum_{j=1}^{n}|z_{\ell_j}| - \frac{\rho}{2}\right) + \frac{|z_{\ell_n}|}{4\theta(1+\omega)(n-1)-1} \\ &\geq -\frac{4\theta(1+\omega)}{(4\theta(1+\omega)n-1)(4\theta(1+\omega)(n-1)-1)}\left(\sum_{j=1}^{n}|z_{\ell_j}| - \frac{\rho}{2}\right) + \frac{s_n(z)}{4\theta(1+\omega)(n-1)-1} \;=\; 0. \end{aligned}$$

We conclude that either $u = 0$ or $\|u\|_1 > 2\theta\rho$. The former case is impossible since $u = 0$ implies $z \in B^1(\rho/2)$, i.e., $\|z\|_1 < \rho/2 < 2\rho\theta(1+\omega)$. Thus, $\|u\|_1 > 2\theta\rho$ and consequently $u = (1+\omega)^{-1}z$ as already noted above.

We finally assume $\|z\|_1 > \rho/2$ and $\|z\|_\infty \leq 2\rho\theta(1+\omega)$. Then certainly $u \neq 0$ since this would imply $z \in B^1(\rho/2)$, i.e., $\|z\|_1 \leq \rho/2$. Moreover, $\|u\|_\infty \leq 2\theta\rho$ since the opposite would result in $z = (1+\omega)u$, i.e., $\|z\|_\infty > 2\rho\theta(1+\omega)$. Hence, by the arguments above there exists

n such that conditions (A.18), (A.19), (A.20) and (A.21) hold. Considering the definition of $h^{(\infty)}_{\theta(1+\omega),\rho}$ we conclude that

$$u = (1+\omega)h^{(\infty)}_{\theta(1+\omega),\rho}(z),$$

in all cases. ■

References

[1] L. Ambrosio and V.M. Tortorelli, *Approximation of functionals depending on jumps by elliptic functionals via* Γ*-convergence.*, Commun. Pure Appl. Math. **43** (1990), no. 8, 999–1036.

[2] S. Anthoine, *Different Wavelet-based Approaches for the Separation of Noisy and Blurred Mixtures of Components. Application to Astrophysical Data.*, Ph.D. thesis, Princeton University, 2005.

[3] D. Baron, M.B. Wakin, M.F. Duarte, S. Sarvotham, and R.G. Baraniuk, *Distributed Compressed Sensing*, preprint (2005).

[4] T. Blumensath, M. Yaghoobi, and M. Davies, *Iterative hard thresholding and ℓ_0 regularisation*, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2007.

[5] E. Candès, J. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math. **59** (2006), no. 8, 1207–1223.

[6] A. Chambolle, *Image segmentation by variational methods: Mumford and Shah functional and the discrete approximations.*, SIAM J. Appl. Math. **55** (1995), no. 3, 827–863.

[7] A. Chambolle, R. A. DeVore, N.-Y. Lee, and B. J. Lucier, *Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage*, IEEE Trans. Image Process. **7** (1998), no. 3, 319–335.

[8] A. Chambolle and P.-L. Lions, *Image recovery via total variation minimization and related problems.*, Numer. Math. **76** (1997), no. 2, 167–188.

[9] O. Christensen, *An Introduction to Frames and Riesz Bases*, Birkhäuser, Boston, 2003.

[10] G. Dal Maso, *An Introduction to* Γ*-Convergence.*, Birkhäuser, Boston, 1993.

[11] I. Daubechies, M. Defrise, and C. DeMol, *An iterative thresholding algorithm for linear inverse problems*, Comm. Pure Appl. Math. **57** (2004), no. 11, 1413–1457.

[12] I. Daubechies and G. Teschke, *Variational image restoration by means of wavelets: Simultaneous decomposition, deblurring, and denoising.*, Appl. Comput. Harmon. Anal. **19** (2005), no. 1, 1–16.

[13] I. Daubechies, G. Teschke, and L. Vese, *Iteratively solving linear inverse problems under general convex constraints*, preprint, 2006, to appear in Inverse Problems and Imaging.

[14] D. L. Donoho, *Superresolution via sparsity constraints.*, SIAM J. Math. Anal. **23** (1992), no. 5, 1309–1331.

[15] ———, *De-noising by soft-thresholding.*, IEEE Trans. Inf. Theory **41** (1995), no. 3, 613–627.

[16] ———, *Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition.*, Appl. Comput. Harmon. Anal. **2** (1995), no. 2, 101–126.

[17] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inf. Theory **52** (2006), no. 4, 1289–1306.

[18] D. L. Donoho and I. M. Johnstone, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika **81** (1994), no. 3, 425–455.

[19] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard, *Wavelet shrinkage: asymptopia?*, J. Roy. Statist. Soc. Ser. B **57** (1995), no. 2, 301–369.

[20] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, *Least angle regression*, Ann. Statist. **32** (2004), no. 2, 407–499.

[21] M. Elad, J.-L. Starck, P. Querre, and D.L. Donoho, *Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)*, Appl. Comput. Harmon. Anal. **19** (2005), 340–358.

[22] M. A. T. Figueiredo and R. D. Nowak, *An EM algorithm for wavelet-based image restoration.*, IEEE Trans. Image Proc. **12** (2003), no. 8, 906–916.

[23] M. Fornasier and F. Pitolli, *Adatpive iterative thresholding algorithms for magnetoencephalography (MEG)*, (2007), preprint.

[24] M. Fornasier and H. Rauhut, *Recovery algorithms for vector valued data with joint sparsity constraints*, preprint (2006), `http://homepage.univie.ac.at/holger.rauhut/joint_sparsity_final.pdf`.

[25] H. Gao and A. G. Bruce, *WaveShrink with firm shrinkage*, Statist. Sinica **7** (1997), no. 4, 855–874.

[26] I. Loris, G. Nolet, I. Daubechies, and F. A. Dahlen, *Tomographic inversion using ℓ_1-norm regularization of wavelet coefficients*, (2006), preprint.

[27] R. Ramlau and G. Teschke, *Tikhonov replacement functionals for iteratively solving nonlinear operator equations.*, Inverse Probl. **21** (2005), no. 5, 1571–1592.

[28] H. Rauhut, *Random sampling of sparse trigonometric polynomials*, Appl. Comput. Harm. Anal. **22** (2007), no. 1, 16–42.

[29] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms.*, Physica D **60** (1992), no. 1-4, 259–268.

[30] J.-L. Starck, E. J. Candès, and D. L. Donoho, *Astronomical image representation by curvelet transform*, Astronomy and Astrophysics **298** (2003), 785–800.

[31] J.-L. Starck, M. K. Nguyen, and F. Murtagh, *Wavelets and curvelets for image deconvolution: a combined approach*, Signal Proc. **83** (2003), 2279–2283.

[32] G. Teschke, *Multi-frames in thresholding iterations for nonlinear operator equations with mixed sparsity constraints*, preprint, 2005.

[33] ———, *Multi-frame representations in linear inverse problems with mixed multi-constraints*, preprint, 2006, to appear in Appl. and Comput. Harm. Anal.

[34] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Roy. Statist. Soc. Ser. B **58** (1996), no. 1, 267–288.

[35] J. Tropp, *Algorithms for simultaneous sparse approximation. Part II: Convex relaxation*, Signal Processing **86** (2006), 589–602.

[36] J. Tropp, K. Herrity, and A. C. Gilbert, *Sparse approximation via iterative thresholding*, Proceedings of the 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2006.

[37] H. Zou, *The Adaptive Lasso and Its Oracle Properties*, J. Amer. Stat. Assoc. **101** (2006), 1418–1429.