Iterative Thresholding Algorithms

Massimo Fornasier* and Holger Rauhut†

July 24, 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 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, frames, variational calculus on sequence spaces, Γ-convergence

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

1 Introduction

This paper addresses the solution of an inverse problem $Af = g$, given possibly noisy data g, where A is a linear bounded operator between Hilbert spaces. When this inverse problem is ill-posed then a regularization mechanism is required [20]. Regularization helps to identify the solution of interest, by taking advantage of a priori knowledge. Recently the imposition of *sparsity constraints* as regularization method has proven to be an effective strategy [1, 9, 27, 28, 33]. This assumes that the solution f has a sparse expansion with respect to a suitable basis or frame $\{\psi_\lambda : \lambda \in \Lambda\}$, i.e., it can be well-approximated by a linear combination of few elements of the prescribed frame. This approach has already proven successful in various applications such as deconvolution and super-resolution problems [12, 14, 32], image recovery and enhancing [10, 19], problems arising in geophysics and biomedical imaging [22, 25], statistical estimation [18, 34], and compressed sensing [2, 4, 15, 29]. It is now well-understood that sparsity can be imposed by minimizing the ℓ_1-regularized functional

$$K_\alpha(u) = \|AFu - g\|^2 + \alpha\|u\|_1, \tag{1.1}$$

*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.

where u corresponds to the coefficient vector of the solution $f = Fu := \sum_\lambda u_\lambda \psi_\lambda$. In the simple case that AF is the identity operator (or at least unitary), the minimizer of K_α can be computed explicitly by *soft-thresholding* the components of g [5, 13, 16, 17]. Due to its variational formulation, its simplicity and effectiveness in noise removal, soft-thresholding proved to be an efficient alternative to total variation minimization [30, 6, 11], which requires instead the solution of a degenerate PDE. In the case of a general linear operator AF the solution cannot be computed explicitly, and several authors have proposed soft thresholded Landweber iterations to approximate it [21, 31, 32, 18]. The convergence of the algorithm was later established in [9].

In the case that f is actually a 'multi-channel' signal consisting of several components (i.e., a vector valued function) we speak of *joint sparsity* if all the components possess a sparse representation and additionally the non-zero (significant) coefficients appear at the same locations [2, 35]. For instance, color images divided into channels (e.g., RGB) can be considered as jointly sparse signals since edges, and hence, significant wavelet or curvelet coefficients, appear at the same locations. Similar concepts appeared recently also in statistical estimation when input variables are grouped together [26, 37].

In [23] we introduced an extension of the functional (1.1) modelling jointly sparse recovery. It uses a weighted ℓ_1-norm of 'interchannel' ℓ_q-norms of the frame coefficients as penalty term. Additionally the weights are chosen adaptively in the sense that they are treated as optimization variables as well, see (2.4) below. We developed and analyzed an iterative minimization algorithm in [23] that alternates between minimizing with respect to the coefficients u (which requires another inner iteration) and with respect to the weights, see (2.8). When investigating the role of the parameters defining the functional in the scalar (single-channel) case we discovered a surprising correspondence between the minimizer of the functional and so-called firm-thresholding [24], and in a special case to hard-thresholding. So as one of the main contributions of this paper, we associate hard-thresholding to the minimizer of a *convex* functional. Further, we provide natural extensions of firm-thresholding operators to the multi-channel case. It is interesting to note that (single-channel) firm-thresholding also arises in so-called iterative refinement algorithms [36]. Similar approaches to adaptive weights appeared as well in statistical estimation, see e.g. [38] and references therein.

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 the (strict) convexity of the functional, we prove such convergence. We note, however, that, unless AF is unitary, these restrictions on the parameters exclude pure iterative hard thresholding. Nevertherless the convergence of iterative hard thresholding to local minimizers of a certain non-convex functional is shown in [3]. Compared with our first alternating algorithm, the new approach clearly has the advantage of providing a single iteration scheme, and we expect that it will have faster convergence in practice; the investigation of this issue is beyond the scope of this paper.

We will also discuss the dependence of the minimizers on the parameters. It is very natural to question how the action of different thresholding operators influences minimizers of the corresponding functionals. Indeed, the minimizers are weakly continuous with respect to the parameters. In particular, our analysis makes explicit the continuity by showing that minimizers of our functional do converge to the minimizers of the ℓ_1-regularized functional

(1.1) for certain limits of the parameters.

The paper is organized as follows. Section 2 introduces notations and our functional. Further, we recall the alternating minimization algorithm in [23] 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 A functional modelling joint sparsity

Let $A : \mathcal{K}' \to \mathcal{H}$ be a linear and bounded operator acting between the separable Hilbert spaces $\mathcal{K}'$ and $\mathcal{H}$. Given data $g \in \mathcal{H}$,

$$Af = g$$

our task is to reconstruct the unknown $f \in \mathcal{K}'$. This problem is possibly ill-conditioned and the data g might be noisy. Hence, regularization is required. Instead of applying classical Tikhonov regularization [20] we will regularize by sparsity constraints [9], or more generally joint sparsity constraints as suggested in [23]. We assume that f is actually a 'multi-channel vector', i.e., $f = (f_1, \dots, f_L)$ with $f_\ell \in \mathcal{K}$, $\ell = 1, \dots, L$ for some Hilbert space $\mathcal{K}$; in other words $\mathcal{K}' = \mathcal{K}^L$. In order to model (joint) sparsity we assume to have a 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

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

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 $u = (u_\lambda)_{\lambda \in \Lambda}$ is contained in the sequence space $\ell_2(\Lambda)$ endowed with the usual ℓ_2-norm. The linear operator $F : \ell_2(\Lambda) \to \mathcal{K}$ is called the *synthesis map* in frame theory [7]. 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.

By using frames the problem of recovering $f_\ell \in \mathcal{K}$ can be restated in terms of frame coefficients in $\ell_2(\Lambda)^L$. To this end we introduce the operator

$$T : \ell_2(\Lambda)^L \to \mathcal{H}, \quad Tu = A\left(Fu^1, Fu^2, \dots, Fu^L\right) = A\left(\sum_{\lambda \in \Lambda} u_\lambda^1 \psi_\lambda, \dots, \sum_{\lambda \in \Lambda} u_\lambda^L \psi_\lambda\right).$$

Then our problem is reformulated as solving the equation

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

for the frame coefficients u. Once a solution $u = (u_\lambda^\ell)$ is determined we obtain a reconstruction of the vectors of interest by means of $f_\ell = Fu^\ell = \sum_\lambda u_\lambda^\ell \psi_\lambda$. The coefficient vector u and the corresponding f are called sparse if u has only a small number of non-zero coefficients. Generalizing slightly this concept we say that u is jointly sparse if all its components u^ℓ, $\ell = 1, \ldots, L$ are sparse, and additionally the support sets for all channels u^ℓ are the same. Hence, all f_ℓ can be represented as $f_\ell = \sum_{\lambda\in\Lambda_0} u_\lambda^\ell \psi_\lambda$ where Λ_0 is small (finite) and coincides for all $\ell = 1, \ldots, L$.

Many types of signals can be well-approximated by sparse ones if the frame is suitably chosen. Joint sparsity naturally occurs for instance in color images, where, e.g., the three color channels RGB can usually be well approximated by a jointly sparse wavelet or curvelet expansion since edges appear at the same locations throughout all channels. The key idea is to incorporate joint sparsity in the regularization of the inverse problem (2.3). In [23] we proposed to work with the functional

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

where $\|\cdot\|_q$ denotes the usual q-norm on $\mathbb{R}^L$, $q \in [1,\infty]$ and $\theta = (\theta_\lambda), \omega = (\omega_\lambda)$ and $\rho = (\rho_\lambda)$ are suitable sequences of positive parameters. The variable u is assumed to be in $\ell_2(\Lambda)^L$ and $v_\lambda \geq 0$ for all $\lambda \in \Lambda$. Observe, that u_λ is a vector in $\mathbb{R}^L$ while v_λ is just a nonnegative 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. As argued in [23] J promotes joint sparsity, i.e., u^* can be expected to be jointly sparse. Further, we note that the functional is even interesting in the monochannel case $L = 1$. Then it just promotes usual sparsity and provides an alternative to ℓ_1-minimization as analyzed in [9]. At this point, it is useful to denote the 'sparsity measure' by

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

which allows to write

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

In [23] 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*

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

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

$$\kappa_q = \begin{cases} L, & q = 1 \\ 1, & q = 2, \\ 1, & q = \infty. \end{cases} \qquad (2.7)$$

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 [23, Proposition 2.1]. □

Usually one has $s_{\min} = 0$ and then (2.6) reduces to the condition already provided in [23]. 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 [23]. Further, we expect that condition (2.6) with suitable κ_q is also sufficient in the general case $q \in [1, \infty]$.

2.2 An algorithm for the minimization of J

In [23] 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

$$\begin{aligned} u^{(n)} &:= \arg\min_{u\in\ell_2(\Lambda)^L} J(u, v^{(n-1)}), \\ v^{(n)} &:= \arg\min_{v\geq 0} J(u^{(n)}, v). \end{aligned} \qquad (2.8)$$

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

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

The minimization of $J(u, v^{(n-1)})$ with respect to u and fixed $v^{(n-1)}$ can be done by a thresholded Landweber iteration similar to the one analyzed in [9]. So let $v = (v_\lambda)_{\lambda\in\Lambda}$ be a fixed positive sequence and $u_{(0)} \in \ell_2(\Lambda)^L$ be some arbitrary initial point and define

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

where

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

and

$$S^{(q)}_v(x) = x - P^{q'}_{v/2}(x), \quad x \in \mathbb{R}^L, \tag{2.12}$$

with $P^{q'}_{v/2}$ denoting the orthogonal projection onto the unit ball of radius $v/2$ in $\mathbb{R}^L$ with respect to the q'-norm where $1/q' + 1/q = 1$. For $q \in \{1, 2, \infty\}$ explicit formulas for $S^q_{v/2}$ are given in [23]. By extending the arguments in [9], we proved that the iteration (2.10) strongly converges to the minimizer of $K(u) = J(u, v)$ under mild conditions on v and ω [23, Proposition 4.9]. Moreover, we showed in [23] that the full algorithm (2.8) indeed converges to the minimizer of the functional J.

3 Relation to Hard and Firm Thresholding

The functional $J = J^{(q)}_{\theta,\rho,\omega}$ depends on several 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 $L = 1$ and the parameter $\omega = 0$ for the moment. (The choice of q becomes clearly irrelevant if $L = 1$). Here the operator T is assumed to be the identity on $\ell_2(\Lambda)$. This leads to the study of the functional

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

Since $s_{\min}(I) = 1$, according to Lemma 2.1 a sufficient (and actually necessary) condition for the 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

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

where

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

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

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

$$u^* = h_{\theta,\rho}(z)$$
$$v^* = \begin{cases} \rho - \frac{1}{2\theta}|u^*|, & |u^*| < 2\theta\rho, \\ 0, & \text{otherwise} , \end{cases}$$

where

$$h_{\theta,\rho}(z) = \begin{cases} 0, & |z| \leq \rho/2, \\ \frac{4\theta}{4\theta-1}\left(z - \operatorname{sign}(z)\frac{\rho}{2}\right), & \rho/2 < |z| \leq 2\theta\rho, \\ z, & |z| > 2\theta\rho. \end{cases} \tag{3.1}$$

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

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

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 anymore.

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 [24], see Figure 1 for a plot. Furthermore, letting $\theta \to \infty$ in the above lemma, we recover the

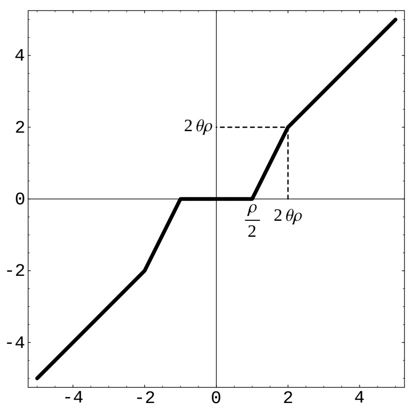

Figure 1: Typical shape of the function $h_{\theta,\rho}$. Here the parameters are $\rho = 2$ and $\theta = 1/2$.

soft-thresholding function,

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

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 $L \ge 1$ with non-trivial parameter ω_λ, but still with T being the identity on $\ell_2(\Lambda)^L$. Then our functional has the form

$$J(u,v) \;=\; J^{(q)}_{\theta,\rho,\omega}(u,v) \;=\; \|u-g\|_2^2 + \sum_{\lambda\in\Lambda}\omega_\lambda\|u_\lambda\|_2^2 + \sum_{\lambda\in\Lambda} v_\lambda\|u_\lambda\|_q + \sum_{\lambda\in\Lambda}\theta_\lambda(\rho_\lambda - v_\lambda)^2 \quad (3.2)$$

with $u \in \ell_2(\Lambda)^L$ and $v_\lambda \ge 0$. By Lemma 2.1 a sufficient (and actually necessary) condition for the convexity of J in the cases $q \in \{1,2,\infty\}$ is

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

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

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

with

$$G^{(q)}_{\theta,\rho,\omega;z}(\tilde u,\tilde v) \;:=\; \|\tilde u - z\|_2^2 + \omega\|\tilde u\|_2^2 + y\|\tilde u\|_q + \theta(\rho-\tilde v)^2, \quad \tilde u \in \mathbb{R}^L, \tilde v \in \mathbb{R}_+. \quad (3.3)$$

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

Before stating the theoretical result let us introduce the following functions for $q = 1, 2, \infty$, respectively.

For $q = 2$, $\theta > 1/4$ and $z \in \mathbb{R}^L$ we define

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

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

1. If $\|z\|_\infty < \rho/2$ then
$$h^{(1)}_{\theta,\rho}(z) \;:=\; 0.$$
2. If $\|z\|_1 \ge 2\theta\rho$ 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, $|z_{\ell_1}| \geq |z_{\ell_2}| \geq \ldots \geq |z_{\ell_L}|$. For $n = 1, \ldots, L$ define

$$t_n(z) := \rho/2 - \frac{\sum_{j=1}^n |z_{\ell_j}| - \frac{n\rho}{2}}{4\theta - n}. \tag{3.4}$$

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

$$|z_{\ell_n}| \geq t_n(z) \tag{3.5}$$

and

$$|z_{\ell_{n+1}}| < t_n(z) \tag{3.6}$$

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

$$\begin{aligned} (h_{\theta,\rho}^{(1)}(z))_{\ell_j} &:= z_{\ell_j} - \text{sign}(z_{\ell_j}) t_n(z), \qquad j = 1, \ldots, n, \\ (h_{\theta,\rho}^{(1)}(z))_{\ell_j} &:= 0, \qquad j = n+1, \ldots, L. \end{aligned}$$

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

1. If $\|z\|_1 < \rho/2$ then

$$h_{\theta,\rho}^{(\infty)}(z) = 0.$$

2. If $\|z\|_\infty \geq 2\theta\rho$ then

$$h_{\theta,\rho}^{(\infty)}(z) = z.$$

3. If $\|z\|_1 \geq \rho/2$ and $\|z\|_\infty < 2\theta\rho$ then we order the coefficients of z by magnitude, $|z_{\ell_1}| \geq |z_{\ell_2}| \geq \ldots \geq |z_{\ell_L}|$. 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 $\{1, \ldots, L\}$ such that $s_m(z) \geq 0$. (Such m exists since $s_L(z) \geq 0$ follows from $\|z\|_1 \geq \rho/2$.) As follows from the proof of the next theorem there exists a unique $n \in \{m, \ldots, L\}$ 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 $n = L$). Then we define the components of $h_{\theta,\rho}^{(\infty)}$ as

$$\begin{aligned} (h_{\theta,\rho}^{(\infty)}(z))_{\ell_j} &:= \text{sign}(z_{\ell_j}) s_n(z), \qquad j = 1, \ldots, n, \\ (h_{\theta,\rho}^{(\infty)}(z))_{\ell_j} &:= z_{\ell_j}, \qquad j = n+1, \ldots, L. \end{aligned}$$

These functions $h^{(q)}_{\theta,\rho}$ provide different generalizations of the firm shrinkage function $h_{\theta,\rho}$ in (3.1) 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 $z \in \mathbb{R}^L$. Assume*

$$(\omega + 1)\theta > \kappa_q/4 \tag{3.7}$$

with κ_q in (2.7) ensuring strict convexity of the function $G^{(q)}_{\theta,\rho,\omega;z}$ in (3.3) by Lemma 2.1. Then the minimizer $(u, v) \in \mathbb{R}^L \times \mathbb{R}_+$ of $G^{(q)}_{\theta,\rho,\omega;z}$ is given by

$$\begin{aligned} u &= (1+\omega)^{-1} h^{(q)}_{\theta(1+\omega),\rho}(z), \\ v &= \begin{cases} \rho - \frac{\|u\|_q}{2\theta}, & \|u\|_q < 2\theta\rho, \\ 0, & \textit{otherwise} . \end{cases} \end{aligned} \tag{3.8}$$

The proof of this theorem is rather long and technical, and therefore postponed to the Appendix. We note that condition (3.7) is required to ensure uniqueness of the minimizer of $G_{\theta,\rho,\omega;z}$. In case of equality in (3.7) 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 the trivial identity operator $T = I$ in (3.2) is clearly given by

$$u^* = H^{(q)}_{\theta,\rho,\omega}(g), \tag{3.9}$$

$$v^* = V^{(q)}_{\theta,\rho}(u^*), \tag{3.10}$$

where

$$\left(H^{(q)}_{\theta,\rho,\omega}(g) \right)_\lambda := (1+\omega_\lambda)^{-1} h^{(q)}_{\theta_\lambda(1+\omega_\lambda),\rho_\lambda}(g_\lambda), \tag{3.11}$$

and

$$\left(V^{(q)}_{\theta,\rho}(u^*) \right)_\lambda := \begin{cases} \rho_\lambda - \frac{1}{2\theta_\lambda}\|u^*_\lambda\|_q, & \|u^*_\lambda\|_q < 2\theta_\lambda\rho_\lambda \\ 0, & \text{otherwise} . \end{cases} \tag{3.12}$$

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 $v = V^{(q)}_{\theta,\rho}(H_{\theta,\rho,\omega}(g))$. Then*

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

Proof. Let (u^*, v^*) be the minimizer of the functional J in (3.2). Then $u^* = H^{(q)}_{\theta,\rho,\omega}(g)$ and $v^* = V^{(q)}_{\theta,\rho}(u)$ by (3.9) and (3.10). Since (u^*, v^*) minimizes $J(u, v)$, we have in particular $u^* = \arg\min_u J(u, v^*)$. By Lemma 4.1 in [23] it holds $u^* = U^{(q)}_{v^*,\omega}(g)$, 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 $z \in \mathbb{R}^L$ and some $v^{(0)} \in \mathbb{R}_+$ define for $n \geq 1$*

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

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

$$J_z(u, v) \;=\; \|u - z\|_2^2 + v\|u\|_q + \theta(\rho - v)^2, \quad u \in \mathbb{R}^L, v \in \mathbb{R}_+, \tag{3.13}$$

is convex. The proposed iteration corresponds precisely to the scheme (2.8) and by Theorem 3.1 in [23] the scheme thus converges. The error estimate follows from Proposition 5.4 in [23]. □

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 (3.13) 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 $L \geq 1$ 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 a non-trivial operator T. In contrast to the algorithm (2.8) analyzed in [23], it consists only of a single iteration rather than an alternating minimization algorithm.

We first need to introduce surrogate functionals similar to the one in [9]. For some additional parameter $a \in \ell_2(\Lambda)^L$ 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 $u^{(0)} \in \ell_2(\Lambda)^L$ 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 (3.11) and (3.12). 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 again 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 [23] the subdifferential of J at (u, v) contains the set

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

where

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

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 proposition 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 [23] 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 $u^{(0)} \in \ell_2(\Lambda)^L$ 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 $u^* \in \ell_2(\Lambda)^L$ 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,*

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

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

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 [23, Lemma 5.2] that $\|P^{q'}_v(x) - P^{q'}_w(x)\| \le K_q|v-w|$ for all $v, w \ge 0$, and $x \in \mathbb{R}^L$, with $K_1 = \sqrt{L}$ 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 (3.12) and distinguishing different cases we obtain

$$|V_{\theta,\rho}^{(q)}(H_{\theta,\rho,\omega}^{(q)}(y))_\lambda - V_{\theta,\rho}^{(q)}(H_{\theta,\rho,\omega}^{(q)}(z))_\lambda| \leq \frac{1}{2\theta_\lambda}\left|\|(H_{\theta,\rho,\omega}^{(q)}(y))_\lambda\|_q - \|(H_{\theta,\rho,\omega}^{(q)}(z))_\lambda\|_q\right|$$
$$\leq \frac{1}{2\theta_\lambda}\|(H_{\theta,\rho,\omega}^{(q)}(y))_\lambda - (H_{\theta,\rho,\omega}^{(q)}(z))_\lambda\|_q \leq \frac{R_q}{2\theta_\lambda}\|(H_{\theta,\rho,\omega}^{(q)}(y))_\lambda - (H_{\theta,\rho,\omega}^{(q)}(z))_\lambda\|_2,$$

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

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

Noting that $K_q R_q = \kappa_q$ we obtain

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

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

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

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,

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

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

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

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. □

REMARK: Using similar techniques as in [9], in particular Opial's theorem, one can show (weak) convergence of the algorithm (4.6) also in the case that condition (4.5) is relaxed to

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

Although a thorough numerical study of the algorithms remains to be done we give some hints how to choose the parameters, in particular, q. The higher q, the more joint sparsity is promoted, in particular $q = \infty$ yields the strongest coupling of channels. However, the thresholding operator is the simplest to compute for $q = 2$, so the latter choice might be a good trade-off between joint sparsity and computation speed. The parameter ρ_λ determines up to which level coefficients 'survive' thresholding; in other words, the higher ρ_λ the sparser the solution. The steepness of the firm-thresholding curve between $\rho_\lambda/2$ and $2\rho_\lambda\theta_\lambda$ is governed by θ_λ, and the closer θ_λ approaches 1/4 the close we get to hard-thresholding (provided $\omega_\lambda = 0$). However, in case of non-invertible T the convexity condition (2.6) requires that we stay strictly away from hard-thresholding. Further, ω_λ should be chosen relatively small compared to ρ_λ and θ_λ. Otherwise, the quadratic term in J is dominating, which is known to promote rather many small coefficients, hence, non-sparse solutions. However, a balanced combination of ℓ_1 and ℓ_2 constraints can nevertheless produce sparse solutions.

5 On Variational Limits

In this section we state that the minimizers of $J = J^{(q)}_{\theta,\rho,\omega}$ vary weakly continously with respect to the parameters. This, in turn, shows that slight changes of parameters do not dramatically alter the computed solution. For the sake of brevity, we limit our analysis to show the interesting case where the minimizers of the functional J weakly converge to minimizers of K_ρ as given in (1.1), for certain limits of the parameters. Precisely the same analysis can be generalized to intermediate cases.

5.1 Approaching soft-thresholding

We 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 [9, 23].

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

$$\omega_\lambda^{(k)} - \frac{1}{4\kappa_{q'}\theta_\lambda^{(k)}} \leq \omega_\lambda^{(k-1)} - \frac{1}{4\kappa_q\theta_\lambda^{(k-1)}} \tag{5.3}$$

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 $\ell_2(\Lambda)^L$ 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 [8] 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 $F(x) \leq \liminf_k F(x_k)$.

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

The following well-known result can be achieved as a direct combination of [8, Proposition 5.7, Theorem 7.8, Corollary 7.20, Corollary 7.24].

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 of Theorem 5.1. First we show that K_ρ is coercive and lower-semicontinuous with respect to the weak topology of $\ell_2(\Lambda)^L$. 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 $K_\rho(u) \leq t$, then $\|u\|_2 \leq C_q t$, which shows that $\{u \in \ell_2, K_\rho(u) \leq t\}$ 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 $X = \{u \in \ell_2, K_\rho(u) \leq C\}$ for a sufficiently large C. Observe that by [8, Proposition 8.7] the space X is indeed metrizable with the weak topology inherited from $\ell_2(\Lambda)^L$.

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 $\|Tu - g\|_{\mathcal{H}} \leq \lim_k \|Tu^{(k)} - g\|_{\mathcal{H}}$. 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 $\mathbb{R}^L$ 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

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

Thus, $F_{(k)}(u) \le F_{(k-1)}(u)$ for all $u \in X$ and $k \ge k_0$. In particular, $F_{(k)} \ge K_\rho$ 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 $C > L$ and a sequence $\theta_\lambda^{(k)}$ such that $\theta_\lambda^{(k)} \geq \frac{(C-1)L}{CL-1}\theta_\lambda^{(k-1)}$ and (5.1) is satisfied, for instance

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

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).

A Appendix

Proof of Theorem 3.2

The proof uses subdifferentials. This requires to formally extend the function $G_{\theta,\rho,\omega;z}$ to $\mathbb{R}^L \times \mathbb{R}$ by setting $G_{\theta,\rho,\omega;z}(u,v) = \infty$ if $v < 0$. In [23] the following characterization was provided.

Lemma A.1. *Let $(u,v) \in \mathbb{R}^L \times \mathbb{R}_+$. Then $(\xi,\eta) \in \mathbb{R}^L \times \mathbb{R}$ is contained in the subdifferential $\partial G^{(q)}_{\theta,\rho,\omega;z}(u,v)$ 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 $\mathbb{R}^L$ is given as follows.

- If $1 < q < \infty$ then

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

  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

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

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

- If $q = \infty$ then

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

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

Proof of Theorem 3.2. First observe that $G_{\theta,\rho,\omega;z}$ is strictly convex, continuous on its domain $\mathbb{R}^L \times \mathbb{R}_+$, and bounded from below, further $G_{\theta,\rho,\omega;z}(u,v) \to \infty$ when $\|u\|_2 + |v| \to \infty$. 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 (3.8).

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 $\|z\|_2 \leq \rho/2$. This proves that $u = 0$ if and only if $\|z\|_2 \leq \rho/2$. 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 $\|u\|_2 \leq 2\theta\rho$, this relation can only hold if $\|z\|_2 \leq 2\theta(1+\omega)\rho$.

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 $\|z\|_2 \geq 2\theta(1+\omega)\rho$. 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 $u \neq 0$ and $\|u\|_1 \leq 2\theta\rho$. 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

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

Thus, we need to determine $\|u\|_1$. Let $\ell \in \{1, \dots, L\}$ and assume $u_\ell \neq 0$. 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 $S = \operatorname{supp}(u) = \{\ell : u_\ell \neq 0\}$ 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

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

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

and

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

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 $|z_{\ell_1}| \geq |z_{\ell_2}| \geq \dots \geq |z_{\ell_L}|$ 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 $n = L$. 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 $n \in \{1, \dots, L\}$. 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 $\|z\|_1 \leq L\|z\|_\infty < L\rho/2 < 2\theta(1+\omega)\rho$ by the assumption $\theta(1+\omega) > L/4$. 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 $n=L$. However, there might exist $n=m<L$ for which (A.10) is satisfied. In this case it suffices to show that condition (A.12) is never satisfied for $n=1,\dots,m$. Indeed, for $n \leq m$ we estimate the right hand side of (A.12) as

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

Here we used the ordering of the $|z_{\ell_j}|$ and $4\theta(1+\omega) > L$. 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 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 (3.4) (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 $\ell_1,\dots,\ell_n$. For simplicity we further assume that entries $z_{\ell_1},\dots,z_{\ell_n}$ are positive (the other cases can be treated similarly by taking into account the corresponding signs). Then the numbers $u_{\ell_1},\dots,u_{\ell_n}$ 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

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

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

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

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

- We solve first the linear problem

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

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

$$t = T(\rho - \frac{t}{2\theta}, z_{\ell_1}, \dots, z_{\ell_n}).$$

So, let us solve the linear problem. To this end we follow the computations in [23, 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 $a_j = \frac{2}{nv} \left(v/2 + (n-1) z_{\ell_j} - \sum_{k \in \{1,\ldots,n\} \setminus \{j\}} z_{\ell_k} \right)$. 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

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

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

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

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

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

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 $|z_{\ell_1}| \geq |z_{\ell_2}| \geq \dots \geq |z_{\ell_L}|$.

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 $n \in \{1, \dots, L\}$ such

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

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

and

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

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

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

Now assume that $\|z\|_1 \leq \rho/2$. Then clearly, there exists no $n \in \{1, \dots, L\}$ 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 (3.7). 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] 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.

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

[3] 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.

[4] 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.

[5] 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.

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

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

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

[9] 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.

[10] 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.

[11] I. Daubechies, G. Teschke, and L. Vese, *Iteratively solving linear inverse problems under general convex constraints*, Inverse Problems and Imaging **1** (2007), 29–46.

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

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

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

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

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

[17] 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.

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

[19] 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.

[20] H.W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems.*, Mathematics and its Applications (Dordrecht). 375. Dordrecht: Kluwer Academic Publishers., 1996.

[21] 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.

[22] M. Fornasier and F. Pitolli, *Adaptive iterative thresholding algorithms for magnetoenceophalography (MEG)*, J. Comput. Appl. Math. (2007), to appear.

[23] M. Fornasier and H. Rauhut, *Recovery algorithms for vector valued data with joint sparsity constraints*, SIAM. J. Numer. Anal. (2007), to appear.

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

[25] I. Loris, G. Nolet, I. Daubechies, and F. A. Dahlen, *Tomographic inversion using l_1-norm regularization of wavelet coefficients*, Geophys. J. Int. **170** (2007), 359–370.

[26] L. Meier, S. van de Geer, and P. Bühlmann, *The group lasso for logistic regression*, preprint (2006).

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

[28] ______, *A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints*, Numer. Math. **104** (2006), no. 2, 177–203.

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

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

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

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

[33] G. Teschke, *Multi-frame representations in linear inverse problems with mixed multi-constraints*, Appl. Comput. Harmon. Anal. **22** (2007), no. 1, 43–60.

[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. Xu and S. Osher, *Iterative regularization and nonlinear inverse scale space applied to wavelet-based denoising*, IEEE Trans. Image Proc. **16** (2007), no. 2, 534–544.

[37] M. Yuan and Y. Lin, *Model selection and estimation in regression with grouped variables*, J. R. Stat. Soc. Ser. B Stat. Methodol. **68** (2006), no. 1, 49–67.

[38] H. Zou, *The adaptive lasso and its oracle properties*, J. Amer. Stat. Assoc. **101** (2006), 1418–1429.