ITERATED HARD SHRINKAGE FOR MINIMIZATION PROBLEMS WITH SPARSITY CONSTRAINTS

KRISTIAN BREDIES AND DIRK A. LORENZ *

Abstract. A new iterative algorithm for the solution of minimization problems in infinite-dimensional Hilbert spaces which involve sparsity constraints in form of ℓ^p-penalties is proposed. In contrast to the well-known algorithm considered by Daubechies, Defrise and De Mol, it uses hard instead of soft shrinkage. It is shown that the hard shrinkage algorithm is a special case of the generalized conditional gradient method. Convergence properties of the generalized conditional gradient method with quadratic discrepancy term are analyzed. This leads to strong convergence of the iterates with convergence rates $\mathcal{O}(n^{-1/2})$ and $\mathcal{O}(\lambda^n)$ for $p = 1$ and $1 < p \leq 2$ respectively. Numerical experiments on image deblurring, backwards heat conduction, and inverse integration are given.

Key words. sparsity constraints, iterated hard shrinkage, generalized conditional gradient method, convergence analysis

AMS subject classifications. 46N10, 49M05, 65K10

1. Introduction.

This article deals with the solution of minimization problems in infinite-dimensional Hilbert spaces which involve so-called sparsity constraints. Sparsity has been found as a powerful tool in several problems in recent years. It has been recognized, that sparsity is an important structure in many applications ranging from image processing to problems from engineering sciences. Throughout the article the following example will be used for illustration: Minimize the functional

$$\Psi(u) = \frac{\|Ku - f\|^2}{2} + \sum_k w_k |\langle u,\, \psi_k\rangle|^p \tag{1.1}$$

where $K : \mathcal{H}_1 \to \mathcal{H}_2$ is an operator between two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, $\{\psi_k\}$ is an orthonormal basis of $\mathcal{H}_1$, $w_k \geq w_0 > 0$ is a weighting sequence, and for the exponent it holds $1 \leq p \leq 2$. In the following we will use the abbreviation $\langle u,\, \psi_k\rangle = u_k$ for the coefficients of u with respect to the basis $\{\psi_k\}$.

Problems of this type arise in different contexts:

Sparse inverse problems [8]. Here K is a compact operator and one aims at solving the equation $Ku = f$. Furthermore one assumes that the right hand side is not known precisely but up to a certain precision $\|f - \bar{f}\| \leq \delta$. Since K is compact the solution of the problem $Ku = f$ is ill posed. A way out is to regularize the inversion of K by prior knowledge. As proved in [8] the minimization of the above functional Ψ provides a regularization which promotes sparsity of the solution in the basis $\{\psi_k\}$.

Image deblurring [1]. Consider an image f which is degraded by blurring and noise, i.e. $f = K\bar{f} + \delta$. A standard Tikhonov regularization with a quadratic penalty $\Phi(u) = |u|^2_{H^s}$ would lead to a smooth minimizer with still blurred edges. A regularization better adapted to the situation of images is the penalization with the total variation $\Phi(u) = |u|_{TV}$ or (better suited for computation) the Besov semi-norm $\Phi(u) = |u|_{B^1_{1,1}}$, while the latter can be expressed precisely as in (1.1) with a wavelet basis $\{\psi_k\}$.

*kbredies@math.uni-bremen.de, dlorenz@math.uni-bremen.de, Center for Industrial Mathematics / Fachbereich 3, University of Bremen, Postfach 33 04 40, 28334 Bremen, Germany

Sparse representations in dictionaries [10]. The minimization problem (1.1) appears as the so-called basis pursuit in the problem of finding sparse representations in dictionaries in the case of finite-dimensional spaces. Assume we have a noisy signal $f \in \mathbf{R}^n$ and seek for an approximation which is composed by a small number of "atoms" $\{\psi_k\}_{k=1,\ldots,N} \in \mathbf{R}^n$. This can be stated as a constrained minimization problem

$$\min_{a \in \mathbf{R}^N} \sum |a_k| \text{ subject to } \|Da - f\|^2 \leq \delta$$

where $D = [\psi_1 \cdots \psi_N] \in \mathbf{R}^{n \times N}$. The unconstrained problem with Lagrange multiplier λ (depending on f and δ)

$$\min_{a \in \mathbf{R}^N} \sum |a_k| + \lambda \|Da - f\|^2$$

has precisely the same form as (1.1). See also [18, 24].

Operator equations with sparse frame expansions [23]. One can drop the assumption that the solution has a sparse representation in a given basis and consider the solution to be sparse in a given frame $\{\eta_k\}$ (see [7] for an introduction to frames). If one wants to solve the equation $Ku = f$ under the assumption that u has a sparse representation in the given frame, i.e. $u = \sum_k a_k \eta_k$ with only a few $a_k \neq 0$, one solves the minimization problem

$$\min_{a \in \ell^2} \frac{\|KFa - f\|^2}{2} + w_0 \sum_k |a_k|$$

($a = (a_k)$ and $Fa = \sum_k a_k \eta_k$).

Problems of type (1.1) are well analyzed in the case of finite-dimensional Hilbert spaces, see e.g. [5, 9, 14, 16, 17, 19, 21]. While in the finite dimensional case there are powerful minimization techniques based on convex programming [5, 3], the infinite-dimensional case is much harder. One well understood algorithm for the minimization of Ψ is the iterated soft shrinkage algorithm introduced independently in [15] and [22]. The algorithm is analyzed in [1, 6, 8] while in [6, 8] the authors also show convergence of the algorithm in the infinite-dimensional setting. In [8] the authors use techniques based on optimization transfer and in [6] the notion of proximal mappings and forward-backward splitting is used. In this paper we utilize a generalization of the conditional gradient method, see [2].

While the question of convergence is answered it is still open how fast the iteration converges in the infinite-dimensional case. Up to our knowledge no convergence rates have been proven for the iterated soft shrinkage. The main contribution of this article is a new minimization algorithm for which convergence rates are proved. Namely we prove that our algorithm converges linearly for $p > 1$ and like $n^{-1/2}$ for $p = 1$.

The article is organized as follows: In §2 we briefly introduce our new algorithm and state the main results. Section 3 is devoted to the analysis of convergence rates of the generalized conditional gradient method for functionals $F + \Phi$ with a smooth part F and a non-smooth part Φ. We consider a special case, adapted to the problem of minimizing (1.1). The application of the results to the special case (1.1) is given in §4 where explicit rates of convergence the new algorithm is derived and the main results on convergence rates are proven. Section 5 presents numerical experiments on the convergence of our algorithm and compares our algorithm and the iterated soft shrinkage.

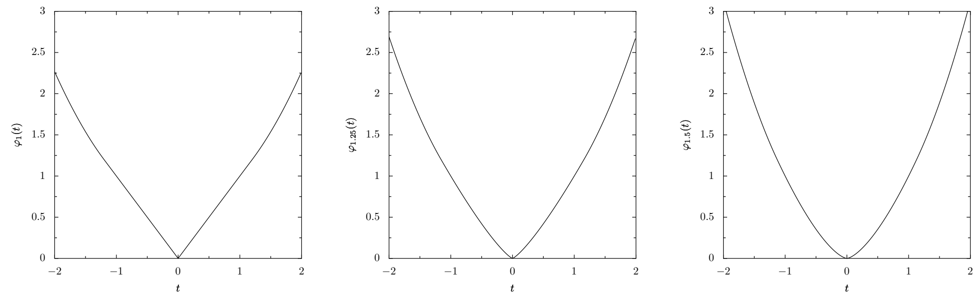

FIG. 2.1. *The function φ_p for $p = 1, 1.25, 1.5$.*

2. An iterative hard shrinkage algorithm. We state the problem of minimizing (1.1), i.e.

$$\min_{u \in \mathcal{H}_1} \ \Psi(u) \ ,$$

with the help of the basis expansion as

$$\min_{u \in \ell^2} \sum_k \frac{(Ku - f)_k^2}{2} + w_k |u_k|^p. \tag{2.1}$$

Remark 2.1. Note that we reformulated the problem. With abuse of notation we used $u = \{u_k\}$, $f = \{f_k\}$ and the operator $\{u_k\} \mapsto K \sum_k u_k \psi_k$ mapping from ℓ^2 to $\mathcal{H}_2$ also denoted by K.

To state our algorithm we introduce the following functions and constants: We denote $S_0 = \left(\frac{\|f\|^2}{2w_0}\right)^{1/p}$ (recall that $0 < w_0 \le w_k$) and the functions

$$\varphi_p(x) = \begin{cases} |x|^p & \text{for } |x| \le S_0 \\ \frac{p}{2S_0^{2-p}} \left(x^2 + (\frac{2}{p} - 1) S_0^2 \right) & \text{for } |x| > S_0 \end{cases} \tag{2.2}$$

and

$$H_p(x) = \begin{cases} \left(\frac{|x|}{p}\right)^{1/(p-1)} \operatorname{sgn}(x) & \text{for } |x| \le p S_0^{p-1} \\ \frac{S_0^{2-p} x}{p} & \text{for } |x| > p S_0^{p-1} \end{cases} \tag{2.3}$$

where we formally set

$$|x|^{\frac{1}{0}} = \begin{cases} 0 & \text{if } |x| \le 1 \\ \infty & \text{if } |x| > 1. \end{cases}$$

Note that φ_p is the usual power for small values and becomes quadratic outside of $[-S_0, S_0]$ in a $\mathcal{C}^1$-way. The function H_p is a kind of shrinkage for small values (remember that $1 \le p \le 2$) and linear outside of $[-S_0, S_0]$. Also note that H_p and φ_p satisfy $H_p(x) \in \partial \varphi_p^{-1}(x)$, a relation which is explained later in §4. See Figure 2.1 and Figure 2.2 for illustrations of φ_p and H_p, respectively.

The minimization algorithm, which turns out to be a special case of the generalized conditional gradient algorithm now reads as follows:

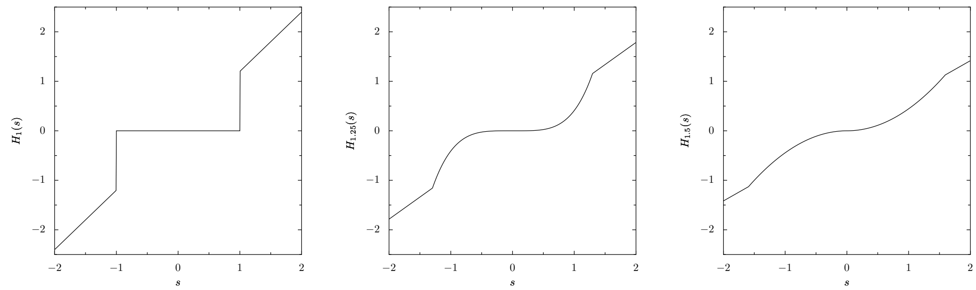

FIG. 2.2. *The function H_p for $p = 1, 1.25, 1.5$.*

ALGORITHM 2.2.

1. **Initialization.** Set $u^0 = 0$ and $n = 0$.
2. **Direction determination.** For $u^n \in \ell^2$ calculate

$$v^n = \mathbf{H}_{p,w}\bigl(-K^*(Ku^n - f)\bigr)$$

where

$$\mathbf{H}_{p,w}\bigl(-K^*(Ku^n - f)\bigr)_k = H_p\Bigl(\frac{-\bigl(K^*(Ku^n - f)\bigr)_k}{w_k}\Bigr) \tag{2.4}$$

with H_p according to (2.3).

3. **Step size determination.** Calculate s_n according to

$$s_n = \min\Bigl\{1, \frac{\sum_{k=1}^{\infty} w_k\bigl(\varphi_p(u_k^n) - \varphi_p(v_k^n)\bigr) + \bigl(K^*(Ku^n - f)\bigr)_k(u_k^n - v_k^n)}{\|K(v^n - u^n)\|^2}\Bigr\} \tag{2.5}$$

whenever the expression makes sense, otherwise let $s_n = 1$.

4. **Iteration.** Set $u^{n+1} = u^n + s_n(v^n - u^n)$, $n := n + 1$ and continue with Step 2.

The main results of this paper now are the convergence of the sequences generated by Algorithm 2.2 and an estimate on the distance to the true minimizer.

THEOREM 2.3. *If $1 < p \leq 2$, then $u^n \to u^*$ to the unique minimizer of* (2.1) *in ℓ^2 with linear convergence speed, i.e.*

$$\|u^n - u^*\| \leq C\lambda^n$$

with a $0 < \lambda < 1$.

If $p = 1$ and K is injective, then $u^n \to u^*$ in ℓ^2 with convergence speed

$$\|u^n - u^*\| \leq Cn^{-1/2}\ .$$

The proof can be divided into two parts: First we examine the convergence of a general minimization algorithm, namely the generalized conditional gradient algorithm (cf. [2]) with discrepancy term $F(u) = \frac{1}{2}\|Ku - f\|^2$ and derive convergence rates for this procedure under certain conditions. We then apply these results to a functional similar to (2.1) and verify that the convergence criteria are satisfied.

3. Convergence analysis of generalized conditional gradient methods. The aim of this section is to provide convergence results for a general descent algorithm which turns out to be Algorithm 2.2 in the special case of the minimization problem (2.1) (up to a modification which does not change the minimizers). Its purpose is to solve the minimization problem

$$\min_{u \in \mathcal{H}_1} \Psi(u) \quad , \quad \Psi(u) = \frac{\|Ku - f\|^2}{2} + \Phi(u) \tag{3.1}$$

in a Hilbert space $\mathcal{H}_1$, with a linear and continuous operator $K : \mathcal{H}_1 \to \mathcal{H}_2$ and some suitable, convex Φ. Note that throughout this section, we will always refer to the problem of general penalty functionals, i.e. Ψ will always denote a functional according to (3.1).

The algorithm is inspired by the generalized conditional gradient method [2] which addresses the minimization of general functionals of the form

$$\Psi(u) = F(u) + \Phi(u)$$

where F is smooth, but non-convex and Φ is convex but possibly non-smooth, resulting in a non-convex and non-smooth Ψ. Here, we consider the special case where $F(u) = \frac{1}{2}\|Ku - f\|^2$, so problem (3.1) is convex, but still potentially non-smooth.

The generalized conditional gradient method applied to (3.1) and an explicit step-size rule gives the following algorithm:

ALGORITHM 3.1.

1. **Initialization.** Set $n = 0$ and choose u^0 such that $\Phi(u^0) < \infty$.
2. **Direction search.** For $n \geq 0$, calculate a minimizer of the approximate problem

$$\min_{v \in \mathcal{H}_1} \langle K^*(Ku^n - f), v\rangle + \Phi(v) \tag{3.2}$$

and denote a solution by v^n.

3. **Step size rule.** Choose the step size s_n according to

$$s_n = \min\left\{1, \frac{\Phi(u^n) - \Phi(v^n) + \langle Ku^n - f, K(u^n - v^n)\rangle}{\|K(v^n - u^n)\|^2}\right\} \tag{3.3}$$

whenever the expression makes sense, otherwise let $s_n = 1$.

4. **Iteration.** Set $u^{n+1} = u^n + s_n(v^n - u^n)$, $n := n + 1$ and continue with Step 2.

In order to apply the algorithm, we have to ensure that the approximate problem (3.2) in Step 2 always has a solution. This is the case if the following conditions are satisfied:

CONDITION 3.2. *Let $\Phi : H \to \mathbf{R} \cup \{\infty\}$ fulfill*

1. Φ *is proper, convex, and lower semi-continuous,*
2. $\partial\Phi$ *is onto with $(\partial\Phi)^{-1}$ bounded, i.e. $\Phi(u)/\|u\| \to \infty$ whenever $\|u\| \to \infty$.*

In the following, we assume that Condition 3.2 on Φ is always satisfied.

Before analyzing the convergence of the algorithm, let us recall equivalent formulations for first-order necessary conditions for solutions of problem (3.1) and introduce some auxiliary functionals which will appear in the following.

PROPOSITION 3.3. *A $u^* \in \mathcal{H}_1$ is a solution of the minimization problem* (3.1) *if and only if one of the following equivalent conditions is satisfied:*

1. $-K^*(Ku^* - f) \in \partial\Phi(u^*)$,
2. $\langle K^*(Ku^* - f),\, u^*\rangle + \Phi(u^*) = \min_{v\in\mathcal{H}_1} \langle K^*(Ku^* - f),\, v\rangle + \Phi(v)$.

Proof. The first statement is just a reformulation of the minimization property $0 \in K^*(Ku^* - f) + \partial\Phi(u^*)$ with the help of subdifferential calculus. The second statement is obviously equivalent to the first by the definition of the subgradient. ∎

Definition 3.4. *For problem* (3.1) *and sequences* $\{u^n\}$ *generated by Algorithm 3.1, introduce*

$$r(u) = \Psi(u) - \min_{v\in\mathcal{H}_1} \Psi(v) \tag{3.4}$$

$$D(u) = \langle K^*(Ku - f),\, u\rangle + \Phi(u) - \Big(\min_{v\in\mathcal{H}_1} \langle K^*(Ku - f),\, v\rangle + \Phi(v)\Big) \;, \tag{3.5}$$

The expressions r_n *and* D_n *will be used as abbreviations for* $r(u^n)$ *and* $D(u^n)$*, respectively.* Note that from the minimization property (3.2) immediately follows that

$$D_n = \Phi(u^n) - \Phi(v^n) + \langle K^*(Ku^n - f),\, u^n - v^n\rangle \;. \tag{3.6}$$

These terms, and especially D, play a central role in the convergence analysis of this algorithm. Since we regard it as a generalization, the ideas utilized in the following are inspired by [12] where the analysis is carried out for the well-known conditional gradient method.

To prove convergence for Algorithm 3.1 we first derive that D serves as an estimate for r, i.e. for the distance to the minimal value of Ψ in (3.1):

Lemma 3.5. *For the functions* D *and* r *according to* (3.4) *and* (3.5)*, respectively, we have the relation* $D \geq r$*. In particular,* $D(u) = 0$ *if and only if* u *is a solution of minimization problem* (3.1).

Proof. Let $u \in \mathcal{H}_1$ be given. Choose a u^* which satisfies

$$\Psi(u^*) = \min_{u\in\mathcal{H}_1} \Psi(u) \;.$$

Since the minimization problem is well-posed (see [13], for example), such an u^* can always be found.

First observe that always

$$\langle K^*(Ku - f),\, u - u^*\rangle \geq \frac{\|Ku - f\|^2}{2} - \frac{\|Ku^* - f\|^2}{2} \;. \tag{3.7}$$

Now

$$\begin{aligned}\langle K^*(Ku - f),\, u\rangle + \Phi(u) - \Big(\min_{v\in\mathcal{H}_1} \langle K^*(Ku - f),\, v\rangle + \Phi(v)\Big)&\\ \geq \langle K^*(Ku - f),\, u - u^*\rangle + \Phi(u) - \Phi(u^*)&\\ \geq (F + \Phi)(u) - (F + \Phi)(u^*) = r(u)&\end{aligned}$$

by the definition of the minimum and (3.7). The characterization $D(u) = 0$ if and only if u is optimal is just a consequence of the the second statement of Proposition 3.3. ∎

Remark 3.6. One immediate consequence is that the step size rule (3.3) always produces $s_n \in [0, 1]$. Moreover, in case the solution is unique, $s_n = 0$ or $v^n = u^n$ if and only if u^n is the solution of the problem: The sufficiency follows directly from the representation of D_n in (3.6), while the necessity can be seen as follows. If

$Ku^n \neq Kv^n$, then, according to (3.3), $s_n = 0$ since $D_n = 0$. The case $Ku^n = Kv^n$ implies with (3.6) that $\Psi(u^n) = \Psi(v^n)$. Consequently, $u^n = v^n$ by uniqueness of solutions.

Remark 3.7. The above algorithm can be interpreted as a modification of the steepest-descent/Landweber algorithm for the minimization of $\frac{1}{2}\|Ku - f\|^2$. Denote by T the (set-valued) solution operator of the minimization problem (3.2).

The steepest-descent algorithm produces iterates $u^{n+1} = u^n + s_n(v^n - u^n)$ according to

$$v^n = -K^*(Ku^n - f) \qquad\qquad s_n = \frac{\langle Ku^n - f,\, K(u^n - v^n)\rangle}{\|K(v^n - u^n)\|^2}\ .$$

In comparison, Algorithm 3.1 also produces in the same manner, with similar directions and step sizes:

$$v^n \in T\bigl(-K^*(Ku^n - f)\bigr)$$
$$s_n = \min\Bigl\{1, \frac{\Phi(u^n) - \Phi(v^n) + \langle Ku^n - f,\, K(u^n - v^n)\rangle}{\|K(v^n - u^n)\|^2}\Bigr\}.$$

Note that in the generalized conditional gradient algorithm, the descent direction of the steepest descent of the quadratic part F is applied to a generally non-linear operator. Likewise, the step size is essentially the one used in the steepest descent algorithm, except for the presence of Φ. Finally, in the iteration step we can only allow convex combinations, therefore it differs with respect to this restriction.

Now for the convergence analysis we note that this generalized conditional gradient algorithm has very convenient descent properties.

LEMMA 3.8. *Algorithm 3.1 produces a sequence $\{u^n\}$ (and $\{v^n\}$), for which the corresponding r_n satisfy*

$$r_{n+1} - r_n \leq \begin{cases} \frac{-r_n}{2} & \text{if } D_n \geq \|K(v^n - u^n)\|^2 \\ \frac{-r_n^2}{2\|K(v^n - u^n)\|^2} & \text{if } D_n \leq \|K(v^n - u^n)\|^2\ . \end{cases} \tag{3.8}$$

Moreover, there exists a $q > 0$ such that

$$r_{n+1} - r_n \leq -q r_n^2$$

for all $n \geq 0$.

Proof. First note that

$$\begin{aligned} &F\bigl(u^n + s_n(v^n - u^n)\bigr) - F(u^n) \\ &\quad = \frac{\|K\bigl(u^n + s_n(v^n - u^n)\bigr) - f\|^2}{2} - \frac{\|Ku^n - f\|^2}{2} \\ &\quad = s_n\langle Ku^n - f,\, K(v^n - u^n)\rangle + \frac{s_n^2\|K(v^n - u^n)\|^2}{2} \end{aligned}$$

and since Φ is convex

$$\Phi\bigl(u^n + s_n(v^n - u^n)\bigr) - \Phi(u^n) \leq s_n\bigl(\Phi(v^n) - \Phi(u^n)\bigr)\ .$$

Putting both together we get

$$\Psi(u^{n+1}) - \Psi(u^n) \leq s_n\underbrace{\bigl(\Phi(v^n) - \Phi(u^n) + \langle K^*(Ku^n - f),\, v^n - u^n\rangle\bigr)}_{=-D_n} + \frac{s_n^2\|K(v^n - u^n)\|^2}{2}\ .$$

We will now make use of D_n according to (3.5) and (3.6). First assume that $D_n \geq \|K(v^n - u^n)\|^2$. Then the step size rule (3.3) yields $s_n = 1$ and it follows

$$r_{n+1} - r_n \leq -D_n + \frac{D_n}{2} \leq -\frac{r_n}{2} \tag{3.9}$$

by Lemma 3.5. In the case where $D_n \leq \|K(v^n - u^n)\|^2$ we have the step size $s_n = D_n / \|K(v^n - u^n)\|^2$, thus

$$r_{n+1} - r_n \leq \frac{-D_n^2}{\|K(v^n - u^n)\|^2} + \frac{D_n^2}{2\|K(v^n - u^n)\|^2} \leq \frac{-r_n^2}{2\|K(v^n - u^n)\|^2} \ , \tag{3.10}$$

again by Lemma 3.5. This proves the first part.

For the second part, remark that in both of the latter cases, the right hand side is non-positive, so it follows $r_{n+1} \leq r_n$ and especially $r_n / r_0 \leq 1$. Hence, we can deduce from (3.9) that if $D_n \geq \|K(v^n - u^n)\|^2$, then

$$r_{n+1} - r_n \leq \frac{-r_n^2}{2r_0} \ .$$

In the other case, we want to derive an estimate for $\|K(v^n - u^n)\|^2$. Since $\{r_n\}$ is bounded and Φ is coercive, there has to be a $C_1 > 0$ such that $\|u^n\| \leq C_1$ for all n. From convex analysis we know that the solution operator of the minimization problem (2.4) in Step 2 is bounded, whenever the property $\Phi(u)/\|u\| \to \infty$ if $\|u\| \to \infty$ is satisfied (see [20], for example). Thus, it follows that $\|v^n\| \leq C_2$ for some constant $C_2 > 0$. This gives the estimate

$$\|K(v^n - u^n)\|^2 \leq \|K\|^2 (C_1 + C_2)^2$$

and, consequently, (3.10) implies

$$r_{n+1} - r_n \leq \frac{-r_n^2}{2\|K\|^2 (C_1 + C_2)^2} \ .$$

Finally, one obtains the desired estimate if one puts the two cases together and sets

$$q = \min \left\{ \frac{1}{2r_0}, \frac{1}{2\|K\|^2 (C_1 + C_2)^2} \right\} \ . \qquad \square$$

Such an estimate immediately implies that the distances to the minimum behave like $\mathcal{O}(n^{-1})$.

LEMMA 3.9. *The distances to the minimum r_n satisfy*

$$r_n \leq C n^{-1}$$

for some $C > 0$ which is independent of n.

Proof. The proof is a widely known trick for the estimation of the distance to the minimum. You can find a similar proof e.g. in [11]. First note that Lemma 3.8 gives $r_n - r_{n+1} \geq q r_n^2$ and in particular $r_{n+1} \leq r_n$ for all $n \geq 0$. Thus,

$$\frac{1}{r_{n+1}} - \frac{1}{r_n} = \frac{r_n - r_{n+1}}{r_{n+1} r_n} \geq \frac{q r_n^2}{r_{n+1} r_n} \geq q > 0$$

and summing up yields

$$\frac{1}{r_n} - \frac{1}{r_0} = \sum_{i=0}^{n-1} \frac{1}{r_{i+1}} - \frac{1}{r_i} \geq q(n-1) \ .$$

Finally, since $q > 0$, we conclude

$$r_n \leq \bigl(q(n-1) + \tfrac{1}{r_0}\bigr)^{-1} \leq Cn^{-1}$$

with a suitably chosen $C > 0$. ∎

A consequence of this lemma is that the sequence $\{u^n\}$ is a minimizing sequence for Ψ. This immediately implies weak convergence of the algorithm by standard arguments, see [13].

THEOREM 3.10. *Each sequence $\{u^n\}$ generated by Algorithm 3.1 possesses a weakly convergent subsequence whose limit is a solution of the minimization problem $\min_{u\in\mathcal{H}_1} \Psi(u)$. On the other hand, each weak accumulation point of $\{u^n\}$ is a solution.*

Additionally, if K is injective or Φ is strictly convex, then the solution u^* of the minimization problem $\min_{u\in\mathcal{H}_1} \Psi(u)$ is unique and each sequence $\{u^n\}$ generated by Algorithm 3.1 converges weakly to u^*.

In many cases, strong convergence can also be established. For this purpose, we introduce a functional which serves as an estimator of r from below: Let u^* be a minimizer and define

$$R(v) = \langle K^*(Ku^* - f),\, v - u^*\rangle + \Phi(v) - \Phi(u^*) \ . \tag{3.11}$$

Note that since u^* is a solution of the problem (3.1), $-K^*(Ku^* - f) \in \partial\Phi(u^*)$ by Proposition 3.3, so R can be interpreted as some kind of Bregman distance at u^* with respect to Φ.

Some basic properties of R can be seen immediately: The second characterization of optimality for u^* as given in Proposition 3.3 implies $R \geq 0$, while from (3.7), one obtains $r \geq R$. Thus, we have the relation

$$D \geq r \geq R \geq 0 \ ,$$

i.e. we are able to estimate the distance to the minimizer by above and below.

Now, the key to proving strong convergence of sequences $\{u^n\}$ generated by Algorithm 3.1 is to examine the growth behavior of R in the neighborhood of u^*. This is done in the following theorem which is also the main result on convergence rates for the generalized conditional gradient method for problems of type (3.1).

THEOREM 3.11. *Let $\{u^n\}$ be a sequence generated by Algorithm 3.1, $u^* \in \mathcal{H}_1$ a minimizer of* (3.1) *and $M \subset \mathcal{H}_1$ a closed subspace with associated orthogonal projection P_M.*

If we have for each $L > 0$

$$\|v - u^*\| \leq L \quad \Rightarrow \quad R(v) \geq c(L)\|P_M(v - u^*)\|^2$$

with some $c(L) > 0$, then $P_M(u^n) \to P_M(u^*)$ in $\mathcal{H}_1$.

If, moreover, $M^\perp$ is finite-dimensional, then there still exists a subsequence of $\{u^n\}$ which converges strongly to a solution. In particular, if K is injective, then $u^n \to u^*$ to the unique solution with convergence speed

$$\|u^n - u^*\| \leq Cn^{-1/2} \ .$$

In the case $M = \mathcal{H}_1$, the minimizer is unique regardless of K and we can improve the convergence speed to

$$\|u^n - u^*\| \leq C\lambda^{n/2} \quad , \quad r_n \leq r_0\lambda^n$$

with some $0 < \lambda < 1$.

Before we give the proof, we need to establish a short functional analytical lemma.

LEMMA 3.12. *Let $K : \mathcal{H}_1 \to \mathcal{H}_2$ be linear, continuous and injective and let $M \subset \mathcal{H}_1$ be a closed subspace with $M^\perp$ finite-dimensional. Then there exists a $C > 0$ such that for each $u \in \mathcal{H}_1$ holds*

$$\|u\| \leq C\big(\|Ku\| + \|P_M u\|\big) \ .$$

Proof. Assume that this is not the case. Then there exists a sequence $\{u^n\} \subset \mathcal{H}_1$ with $\|u^n\| = 1$ and $\|Ku^n\| + \|P_M u^n\| \leq \frac{1}{n}$. It is immediate that there is a subsequence which converges weakly to some $u \in \mathcal{H}_1$. Moreover, $Ku^n \to 0 = Ku$ as well as $P_M u^n \to 0 = P_M u$ and since $M^\perp$ is finite-dimensional, $P_{M^\perp} u^n \to P_{M^\perp} u$. In particular, $u^n \to u$, so $\|u\| = 1$ and $Ku = 0$, a contradiction to the injectivity of K. □

Proof of Theorem 3.11. From Lemma 3.9 we know that the distances r_n converge to zero with estimate $r_n \leq C_1 n^{-1}$. It is also clear that $\|u^n\| \leq C_2$ for some $C_2 > 0$, so we can find an $L > 0$ such that $\|u^n - u^*\| \leq L$. By assumption and convexity of F,

$$r_n \geq \Phi(u^n) - \Phi(u^*) + \langle K^*(Ku^* - f),\, u^n - u^*\rangle \geq c(L)\|P_M(u^n - u^*)\|^2 \tag{3.12}$$

which implies the convergence $P_M(u^n) \to P_M(u^*)$ with rate

$$\|P_M(u^n) - P_M(u^*)\| \leq \sqrt{\frac{C_1}{c(L)}} n^{-1/2} \ .$$

From Theorem 3.10 we know that there is a weakly convergent subsequence of u^n which converges to a solution. Denote this subsequence also by u^n and its weak limit by u^{**}. If $M^\perp$ is finite-dimensional, it follows $P_{M^\perp}(u^n) \to P_{M^\perp}(u^{**})$. By above, $P_M(u^n) \to P_M(u^*)$ and in particular $P_M(u^{**}) = P_M(u^*)$. So it follows

$$u^n = P_M(u^n) + P_{M^\perp}(u^n) \to P_M(u^{**}) + P_{M^\perp}(u^{**}) = u^{**} \ .$$

The convergence statement in case of uniqueness of the solution then follows from the usual subsequence argument.

Now assume that K is injective. According to Lemma 3.12 we can find a $C_3 > 0$ such that for each $u \in \mathcal{H}_1$ holds

$$\|u\| \leq C_3\big(\|Ku\| + \|P_M u\|\big) \ . \tag{3.13}$$

Since u^* is optimal, $-K^*(Ku^* - f) \in \partial\Phi(u^*)$ (again by Proposition 3.3) meaning that

$$\begin{aligned} r_n &= \frac{\|Ku^n - f\|^2}{2} + \Phi(u^n) - \frac{\|Ku^* - f\|^2}{2} - \Phi(u^*) \\ &\geq \frac{\|Ku^n - f\|^2 - \|Ku^* - f\|^2 - 2\langle Ku^* - f,\, K(u^n - u^*)\rangle}{2} \\ &= \frac{\|K(u^n - u^*)\|^2}{2} \ . \end{aligned}$$

Together with (3.13) follows

$$\begin{aligned}\|u^n - u^*\|^2 &\leq 2C_3^2\big(\|K(u^n - u^*)\|^2 + \|P_M(u^n - u^*)\|^2\big) \\ &\leq 2C_1C_3^2\big(2 + c(L)^{-1}\big)n^{-1}\end{aligned}$$

which proves the asserted convergence rate $n^{-1/2}$.

For the remaining statement, note that if $M = \mathcal{H}_1$, then the assumptions imply that $R(v) > 0$ for $v \neq u^*$. But since $R(v) = 0$ for each solution v, the solution u^* has to be unique. Now to prove the linear convergence speed, we first want to show that $r_{n+1} \leq \lambda r_n$ for some $0 < \lambda < 1$. Note that (3.8) in Lemma 3.8 already gives $r_{n+1} \leq \frac{1}{2} r_n$ if $D_n \geq \|K(v^n - u^n)\|^2$, hence we are interested in obtaining an estimate for the term $-r_n^2/(2\|K(v^n - u^n)\|^2)$. According to (3.12) with $M = \mathcal{H}_1$ and (3.8), we have, for $D_n \leq \|K(v^n - u^n)\|^2$, that

$$r_{n+1} \leq r_n\Big(1 - \frac{r_n}{2\|K(v^n - u^n)\|^2}\Big) \leq r_n\Big(1 - \frac{c(L)\|u^n - u^*\|^2}{2\|K(v^n - u^n)\|^2}\Big) \ . \tag{3.14}$$

We aim at showing that the solution operator T of the minimization problem (3.2) is locally Lipschitz continuous in u^* in some sense. Choose a $u \in \mathcal{H}_1$ with $\|u - u^*\| \leq L$ and denote by $v \in T(u)$ a solution of (3.2). Since v solves the minimization problem, it holds

$$\Phi(v) - \Phi(u^*) + \langle K^*(Ku - f),\, v - u^*\rangle \leq 0$$

thus

$$\begin{aligned}\|K\|^2\|u - u^*\|\,\|v - u^*\| &\geq \langle K^*K(u^* - u),\, v - u^*\rangle \\ &\geq \Phi(v) - \Phi(u^*) + \langle K^*(Ku^* - f),\, v - u^*\rangle = R(v) \geq c(L)\|v - u^*\|^2 \ .\end{aligned}$$

It follows that

$$\|v - u^*\| \leq \frac{\|K\|^2}{c(L)}\|u - u^*\| \ ,$$

which establishes some kind of Lipschitz continuity in u^*. Now we can estimate

$$\begin{aligned}2\|K(v^n - u^n)\|^2 &\leq 2\|K\|^2(\|u^n - u^*\| + \|v^n - u^*\|)^2 \\ &\leq 2\|K\|^2\Big(1 + \frac{\|K\|^2}{c(L)}\Big)^2\|u^n - u^*\|^2 = C_4\|u^n - u^*\|^2 \ .\end{aligned}$$

Plugging this into (3.14) gives

$$r_{n+1} \leq r_n\Big(1 - \frac{c(L)\|u^n - u^*\|^2}{2\|K(v^n - u^n)\|^2}\Big) \leq r_n\Big(1 - \frac{c(L)}{C_4}\Big)$$

if $D_n \leq \|K(v^n - u^n)\|^2$. Choosing a suitable $\frac{1}{2} \leq \lambda < 1$ then gives, for all $n \geq 0$,

$$r_{n+1} \leq \lambda r_n \quad \Rightarrow \quad r_n \leq \lambda^n r_0 \ .$$

Finally, the estimate (3.12) on the norm yields

$$\|u^n - u^*\| \leq C\lambda^{n/2}$$

with $C = \sqrt{r_0/c(L)}$, which proves the linear convergence speed. □

4. Convergence rates for iterated hard shrinkage. In this section, we will show how the generalized conditional gradient method can be used to compute minimizers for (1.1) (or, equivalently, solve (2.1)) and prove the convergence rates of Theorem 2.3. In particular, we apply the generalized conditional gradient method to the modified problem

$$\min_{u \in \mathcal{H}_1} \tilde{\Psi}(u), \qquad \text{where} \qquad \tilde{\Psi}(u) := \sum_k \frac{(Ku-f)_k^2}{2} + w_k \varphi_p(u_k) \tag{4.1}$$

(see also (2.2)) which turns out to be exactly Algorithm 2.2. It is proven that (4.1) admits exactly the same minimizers as (2.1) and the convergence rates then follow from applying Theorem 3.11. Finally, we conclude with some remarks on the algorithm. In the following, we split the functional $\tilde{\Psi} = F + \Phi$ with

$$F(u) = \frac{1}{2}\|Ku - f\|^2 \quad , \quad \Phi(u) = \sum_{k=1}^{\infty} w_k \varphi_p(u_k)$$

in $\mathcal{H}_1 = \ell^2$, with weights $0 < w_0 \leq w_k < \infty$ and φ_p according to (2.2), resulting in a modification of problem (2.1). Our modification is based on the following simple observation: It is clear that for the unmodified Ψ, the objective functional of (2.1), holds $\Psi(u^*) \leq \Psi(0) = \frac{1}{2}\|f\|^2$ for each minimizer u^*. We show in the following (Proposition 4.1) that Ψ fulfills the coercivity condition $\|u\|^p \leq \frac{1}{w_0}\Psi(u)$, so we know $\|u^*\| \leq \bigl(\frac{\|f\|^2}{2w_0}\bigr)^{1/p}$. Hence the minimizers of Ψ do not change if we change the functional according to (4.1). This is made precise in the following proposition and remark:

PROPOSITION 4.1. *Let problem* (2.1) *be given for a fixed linear, continuous* $K : \mathcal{H}_1 \to \mathcal{H}_2$ *and* $f \in \mathcal{H}_2$, $1 \leq p \leq 2$. *Then all minimizers* u^* *satisfy*

$$\|u^*\| \leq S_0 \quad , \quad S_0 = \Bigl(\frac{\|f\|^2}{2w_0}\Bigr)^{1/p} .$$

Consequently, the minimizers of

$$\min_{u \in \mathcal{H}_1} \tilde{\Psi}(u) \quad , \quad \tilde{\Psi}(u) = \frac{\|Ku - f\|^2}{2} + \sum_{k=1}^{\infty} w_k \varphi_p(u_k)$$

coincide with the minimizers of (2.1) *whenever* φ_p *is chosen such that*

$$\varphi_p(t) \begin{cases} = |t|^p & \text{for } |t| \leq S_0 \\ \geq |t|^p & \text{for } |t| > S_0 \ . \end{cases}$$

Proof. Observe that, for $u \neq 0$,

$$1 = \sum_{k=1}^{\infty} \Bigl(\frac{|u_k|}{\|u\|}\Bigr)^2 \leq \sum_{k=1}^{\infty} \Bigl(\frac{|u_k|}{\|u\|}\Bigr)^p \quad \Rightarrow \quad \|u\|^p \leq \sum_{k=1}^{\infty} |u_k|^p$$

hence the estimate follows from

$$\|u^*\|^p \leq \frac{1}{w_0} \sum_{k=1}^{\infty} w_k |u_k^*|^p \leq \frac{(F+\Phi)(u^*)}{w_0} \leq \frac{\Psi(0)}{w_0} = \frac{\|f\|^2}{2w_0} = S_0^p \ .$$

Further note that $\Psi(u) \leq \tilde{\Psi}(u)$ with equality if $\|u\| \leq S_0$. If u^* is a minimizer of Ψ, then we have

$$\tilde{\Psi}(u^*) = \Psi(u^*) \leq \Psi(u) \leq \tilde{\Psi}(u)$$

for all $u \in \mathcal{H}_1$. Thus, u^* is also in minimizer for $\tilde{\Psi}$. On the other hand, if u^* is not a minimizer for Ψ, then there exists a $u \in \mathcal{H}_1$ with $\|u\| \leq S_0$ such that

$$\tilde{\Psi}(u) = \Psi(u) < \Psi(u^*) \leq \tilde{\Psi}(u^*)$$

meaning that u^* is also not a minimizer for $\tilde{\Psi}$. ◻

Remark 4.2. Let us remark that all φ_p as defined in (2.2) indeed fulfill $\varphi_p(t) = |t|^p$ for $|t| \leq S_0$ and $\varphi_p(t) \geq |t|^p$ for all $|t| \geq S_0$. The latter follows from $\varphi_p(\pm S_0) = |t|^p$ and a comparison of the derivatives, i.e.

$$|pt^{p-1}| \leq pS_0^{p-2}|t|$$

for $|t| \geq S_0$.

Remark 4.3. We remark on the possibility of normalizing the constant S_0. For given f and w_k we have $S_0 = \bigl(\frac{\|f\|^2}{2w_0}\bigr)^{1/p}$ and define $K^N = S_0 K$, $w_k^N = S_0^p w_k$. With the substitution $u^N = S_0^{-1}u$ problem (2.1) becomes

$$\min_{u^N \in \mathcal{H}_1} \frac{\|K^N u^N - f\|^2}{2} + \sum_{k=1}^{\infty} w_k^N |u_k^N|^p.$$

This results in a constant $S_0^N = \bigl(\frac{\|f\|^2}{2w_0^N}\bigr)^{1/p} = 1$.

In order to apply the convergence results of the previous section, we have to verify that Algorithm 2.2 corresponds to Algorithm 3.1 in the case of (4.1). This will we done in the following. First, we check that the algorithm is indeed applicable, i.e. we show that the functional Φ meets Condition 3.2.

LEMMA 4.4. *Let $1 \leq p \leq 2$ and $\varphi : \mathbf{R} \to \mathbf{R} \cup \{\infty\}$ with $\varphi(0) = 0$ convex, lower semi-continuous, and such that $\varphi(t)/|t| \to \infty$ if $|t| \to \infty$ as well as $\varphi(t) \geq |t|^p$. Then*

$$\Phi(u) = \sum_{k=1}^{\infty} w_k \varphi(u_k)$$

is proper, convex, lower semi-continuous and fulfills $\Phi(u)/\|u\| \to \infty$ when $\|u\| \to \infty$.

Proof. To see that Φ is proper, convex and lower semi-continuous, we refer to the standard literature on convex analysis [13].

To establish the desired coercivity, suppose the opposite, i.e. that there is a sequence with $\|u^n\| \to \infty$ and $\Phi(u_n)/\|u_n\| \leq C_1$ for a $C_1 > 0$. If there exists a $C_2 > 0$ such that $|u_k^n| \leq C_2$ for all $k, n \geq 1$, then there follows, with the help of $|u_k^n|^2 \leq C_2^{2-p}|u_k^n|^p$ as well as $|t|^p \leq \varphi(t)$,

$$\|u^n\|^2 = \sum_{k=1}^{\infty} |u_k^n|^2 \leq \frac{C_2^{2-p}}{w_0} \sum_{k=1}^{\infty} w_k |u_k^n|^p \leq \frac{C_2^{2-p}}{w_0} \sum_{k=1}^{\infty} w_k \varphi(u_k^n) = \frac{C_2^{2-p}}{w_0} \Phi(u^n) \ .$$

Dividing by $\|u^n\|$ yields $\Phi(u^n)/\|u^n\| \to \infty$, a contradiction.

Likewise, if $|u_k^n|$ is not bounded for all $k, n \geq 1$, there exist sequences k_l, n_l such that $|u_{k_l}^{n_l}| \to \infty$ as $l \to \infty$. Since $|u_{k_l}^{n_l}| \leq \|u^{n_l}\| < \infty$, there cannot be infinitely many

k_l for which n_l is constant, thus $n_l \to \infty$. One can furthermore assume that n_l is strictly monotone increasing. Of course, for the corresponding subsequence u^{n_l} there also holds $\Phi(u^{n_l})/\|u^{n_l}\| \leq C_1$, but $\varphi(u^{n_l}_{k_l})/|u^{n_l}_{k_l}| \to \infty$ implies

$$\frac{\Phi(u^{n_l})}{\|u^{n_l}\|} \geq w_0 \varphi(u^{n_l}_{k_l}) \to \infty \quad \text{for} \quad l \to \infty ,$$

which is again a contradiction. ∎

It is not hard to verify that all φ_p according to (2.2) satisfy the requirements of the lemma. In particular, the property $\varphi_p(t)/|t| \to \infty$ whenever $|t| \to \infty$ follows from the quadratic extension outside of $[-S_0, S_0]$.

The next step is to observe that Algorithm 3.1 for the modified functional (4.1) is given by Algorithm 2.2.

PROPOSITION 4.5. *In the situation of Proposition 4.1, the generalized conditional gradient method (Algorithm 3.1) for solving the minimization problem* (4.1) *is realized by Algorithm 2.2.*

Proof. The proof is given by analyzing the steps of Algorithm 3.1 and comparing them with Algorithm 2.2.

Regarding Step 1, choosing $u^0 = 0$ as initialization is feasible since always $\Phi(0) = 0$. The direction search in Step 2 amounts to solving the minimization problem

$$\min_{v \in \mathcal{H}_1} \sum_{k=1}^{\infty} \big(K^*(Ku - f)\big)_k (v_k - u_k) + w_k \varphi_p(u_k) \tag{4.2}$$

which can be done pointwise for each $k \geq 1$. This involves the solution of

$$\min_{t \in \mathbf{R}} \; st + \bar{w}\varphi_p(t)$$

for given $s, t \in \mathbf{R}$ and $\bar{w} > 0$, for which the solution can be derived as follows: First note that $t \in \mathbf{R}$ is a minimizer if and only if $-\frac{s}{\bar{w}} \in \partial\varphi_p(t)$. For $1 < p \leq 2$, this subgradient reads as

$$\partial\varphi_p(t) = \begin{cases} \{p \operatorname{sgn}(t)|t|^{p-1}\} & \text{for } |t| \leq S_0 \\ \{\frac{p}{S_0^{2-p}} t\} & \text{for } |t| > S_0 \end{cases}$$

while

$$\partial\varphi_1(t) = \begin{cases} [-1, 1] & \text{for } t = 0 \\ \{\operatorname{sgn}(t)\} & \text{for } |t| \leq S_0 \\ \{\frac{t}{S_0}\} & \text{for } |t| > S_0 \ . \end{cases}$$

A solution t is then given by $t \in (\partial\varphi)^{-1}(-\frac{s}{\bar{w}})$, which necessarily means, for $1 < p \leq 2$, that

$$t = \begin{cases} \operatorname{sgn}\big(-\frac{s}{\bar{w}p}\big)\big|-\frac{s}{\bar{w}p}\big|^{1/(p-1)} & \text{for } \big|-\frac{s}{\bar{w}}\big| \leq pS_0^{p-1} \\ -\frac{sS_0^{p-2}}{\bar{w}p} & \text{for } \big|-\frac{s}{\bar{w}}\big| > pS_0^{p-1} \end{cases}$$

where the breaking points $\pm pS_0^{p-1}$ are of course given by the values of $\partial\varphi_p(\pm S_0)$. For $p = 1$ a similar situation occurs:

$$t \in \begin{cases} \{0\} & \text{for } \big|-\frac{s}{\bar{w}}\big| < 1 \\ \operatorname{sgn}\big(-\frac{s}{\bar{w}}\big)[0, S_0] & \text{for } \big|-\frac{s}{\bar{w}}\big| = 1 \\ \big\{-\frac{sS_0}{\bar{w}}\big\} & \text{for } \big|-\frac{s}{\bar{w}}\big| > 1 \ . \end{cases}$$

Since the algorithm does not demand a particular solution, we can choose $t = 0$ if $\left|-\frac{s}{\bar{w}}\right| = 1$. Putting the cases $p = 1$ and $1 < p \leq 2$ together gives that the pointwise minimization problem is indeed solved by

$$t = H_p\Big(-\frac{s}{\bar{w}}\Big)$$

with H_p according to (2.3). Consequently, the operator

$$v = \mathbf{H}_{p,w}\big(-K^*(Ku - f)\big)$$

given by (2.4) yields a solution of (4.2).

Finally, one can easily convince oneself that Step 3 and 4 in Algorithm 2.2 is exactly corresponding to Step 3 and 4 in Algorithm 3.1 with the particular choice of Φ. ∎

Remark 4.6. As Proposition 4.1 and Lemma 4.4 show, it is not necessary to modify the functional Φ in the case $1 < p \leq 2$. But if we apply Algorithm 3.1 to the unmodified functional, we have to evaluate $|s|^{1/(p-1)}$ for possibly great $|s|$. This might lead to numerical problems since $p \approx 1$ leads to high powers and the available range of numbers may be left.

Moreover, it is also not necessary to take a quadratic extension outside of $[-S_0, S_0]$ as done in (2.2). In fact, an arbitrary function φ satisfying the conditions of Proposition 4.1 and Lemma 4.4 is possible. The choice of φ however is reflected in the algorithm when $(\partial\varphi)^{-1}$ is computed. The quadratic extension in (2.2) leads to the linear sections in (2.3) which are easy to compute.

We want to apply the convergence results of Theorem 3.11. For establishing the estimates of the type $R(v) \geq c(L)\|v - u^*\|^2$ we need the following elementary result:

LEMMA 4.7. *Let $1 < p \leq 2$. For each $C_1 > 0$ and $L > 0$ there exists a $c_1(L, C_1) > 0$ such that*

$$|t|^p - |s|^p - p\,\mathrm{sgn}(s)|s|^{p-1}(t - s) \geq c_1(L, C_1)(t - s)^2$$

for all $|s| \leq C_1$ and $|t - s| \leq L$.

Proof. First note that $\varphi(t) = |t|^p$ is twice differentiable with the exception that the second derivative does not exist in 0:

$$\varphi'(t) = p\,\mathrm{sgn}(t)|t|^{p-1} \quad , \quad \varphi''(t) = p(p-1)|t|^{p-2} \ .$$

Since φ'' is locally integrable, the Taylor formula

$$\begin{aligned} \varphi(t) &= \varphi(s) + \varphi'(s)(t - s) + \int_s^t \varphi''(\tau)(t - \tau)\,\mathrm{d}\tau \\ \Rightarrow \quad |t|^p &= |s|^p + p\,\mathrm{sgn}(s)|s|^{p-1}(t - s) + p(p-1)\int_s^t |\tau|^{p-2}(t - \tau)\,\mathrm{d}\tau \end{aligned}$$

still holds. By assumption $|s|, |t| \leq C_1 + L$, hence $|\tau|^{p-2} \geq |C_1 + L|^{p-2}$ for $|\tau| \leq C_1 + L$, showing that

$$\begin{aligned} |t|^p &\geq |s|^p + p\,\mathrm{sgn}(s)|s|^{p-1}(t - s) + p(p-1)(C_1 + L)^{p-2}\int_s^t (t - \tau)\,\mathrm{d}\tau \\ &\geq |s|^p + p\,\mathrm{sgn}(s)|s|^{p-1}(t - s) + c_1(L, C_1)(t - s)^2 \end{aligned}$$

with a suitably chosen $c_1(L, C_1) > 0$. □

LEMMA 4.8. *Denote by u^* a solution of the minimization problem* (4.1) *and by φ_p modified functionals meeting the requirements of Proposition 4.1 and Lemma 4.4. Consider the associated functional R according to* (3.11).

If $1 < p \leq 2$, then for each $L > 0$ there exists a $c(L) > 0$ such that

$$\|v - u^*\| \leq L \quad \Rightarrow \quad R(v) \geq c(L)\|v - u^*\|^2 \ .$$

If $p = 1$ then there exists a closed subspace $M \subset H$ with $M^\perp$ finite-dimensional such that for each $L > 0$ there exists a $c(L) > 0$ with

$$\|v - u^*\| \leq L \quad \Rightarrow \quad R(v) \geq c(L)\|P_M(v - u^*)\|^2 \ .$$

Proof. First consider the case $1 < p \leq 2$. If we have a minimizer u^* then $-K^*(Ku^* - f) \in \partial\Phi(u^*)$. The functional Φ is defined as a pointwise sum, thus, by standard arguments from convex analysis,

$$\big(-K^*(Ku^* - f)\big)_k = w_k p \operatorname{sgn}(u_k^*)|u_k^*|^{p-1}$$

for each $k \geq 1$.

From Proposition 4.1 we know that $|u_k^*| \leq S_0$ and applying Lemma 4.7 with $C_1 = S_0$ and an arbitrary $L > 0$ gives

$$\begin{aligned} w_k\big(\varphi_p(t) - \varphi_p(s) - p \operatorname{sgn}(s)|s|^{p-1}\big) \\ \geq w_0\big(|t|^p - |s|^p - p\operatorname{sgn}(s)|s|^{p-1}(t-s)\big) \geq w_0 c_1(L)(t-s)^2 \end{aligned}$$

for each $|s| \leq C_1$, $|s - t| \leq L$, remembering that $\varphi_p(s) = |s|^p$ for $|s| \leq C_1$ and $\varphi_p(t) \geq |t|^p$. Hence, if $\|v - u^*\| \leq L$,

$$\begin{aligned} R(v) = \sum_{k=1}^{\infty} w_k\big(\varphi_p(v_k) - \varphi_p(u_k^*) - p\operatorname{sgn}(u_k^*)|u_k^*|^{p-1}\big) \\ \geq w_0 c_1(L) \sum_{k=1}^{\infty} |v_k - u_k^*|^2 = c(L)\|v - u^*\|^2 \ . \end{aligned}$$

This proves the desired statement for $1 < p \leq 2$.

Now let $p = 1$ and u^* be a minimizer. Then we know, analogously to the above, that

$$\big(-K^*(Ku^* - f)\big)_k \in w_k \partial\varphi_1(u_k^*)$$

for each $k \geq 1$. Since $\xi = -K^*(Ku^* - f) \in \ell^2$ we have $\xi_k \to 0$ for $k \to \infty$. Hence, we can choose a k_0 such that $|\xi_k| \leq \frac{w_0}{2}$ for $k \geq k_0$. Observe that $\partial\varphi_1$ is monotone and coincides with $\operatorname{sgn}(\,\cdot\,)$ in a neighborhood of 0 with $\partial\varphi_1(0) = [-1, 1]$, so $u_k^* = 0$ for $k \geq k_0$ since the opposite leads to a contradiction. Thus,

$$R(v) = \sum_{k=1}^{\infty} w_k\big(\varphi_1(v_k) - \varphi_1(u_k^*)\big) - \xi_k(v_k - u_k^*) \geq \sum_{k=k_0}^{\infty} w_k\varphi_1(v_k) - \xi_k v_k \ .$$

Due to the construction of φ_1 we can estimate $|t| \leq \varphi_1(t)$ which further leads to

$$R(v) \geq \sum_{k=k_0}^{\infty} w_k |v_k| - v_k \xi_k \geq w_0 \sum_{k=k_0}^{\infty} \frac{|v_k|}{2} = \frac{w_0}{2} \sum_{k=k_0}^{\infty} |v_k - u_k^*| \ .$$

Define

$$M = \{u \in \ell^2 \ : \ u_k = 0 \text{ for } k < k_0\}$$

which is clearly a closed subspace of ℓ^2 with finite-dimensional complement. Choose a v with $\|v - u^*\| \leq L$, then $(v_k - u_k)^2 \leq L^{-1} |v_k - u_k|$ so with $c(L) = w_0/(2L)$ we finally have

$$R(v) \geq c(L) \|P_M(v - u^*)\|^2 \ . \qquad \square$$

Collecting the results of this section, we are able to apply Theorem 3.11 which finally gives our main convergence result for the iterated hard shrinkage procedure in Algorithm 2.2:

THEOREM 4.9. *Consider minimization problem* (2.1)*. If* $1 < p \leq 2$*, then Algorithm 2.2 produces a sequence* $\{u^n\}$ *which converges linearly to the unique minimizer* u^**, i.e.*

$$\|u^n - u^*\| \leq C \lambda^n$$

for a $0 \leq \lambda < 1$*.*

If $p = 1$ *and* K *is injective, then the descent algorithm produces a sequence* $\{u^n\}$ *which converges to the unique minimizer* u^* *in norm with speed*

$$\|u^* - u^n\| \leq C n^{-1/2} \ .$$

Proof. First note that by Proposition 4.1 and Remark 4.2, the minimization problem (2.1) is equivalent to (4.1) and the latter can be solved instead. Additionally, one knows that (2.1) is well-posed, so at least one optimal solution $u^* \in \mathcal{H}_1$ exists. As already remarked, Lemma 4.4 ensures that Condition 3.2 on the (modified) regularization functional Φ is fulfilled, thus Algorithm 3.1 can be applied which yields Algorithm 2.2 as a special case, see Proposition 4.5. Hence, Theorem 3.11 is applicable if the corresponding prerequisites hold.

In the case $1 < p \leq 2$, Lemma 4.8 gives the required estimate on R (associated with the above u^*) where $M = \mathcal{H}_1$. This implies the linear convergence speed by Theorem 3.11. In the case $p = 1$ and K injective, one can use the estimate for closed subspace $M \subset \mathcal{H}_1$ provided by the second part of Lemma 4.8 to obtain the desired convergence behavior, again with Theorem 3.11. $\square$

Remark 4.10. The iterative hard shrinkage algorithm (Algorithm 2.2) is very simple and hence easy to implement. We just need the functions φ_p and H_p available (which can be implemented explicitly) and of course the application of the operators K and K^*. In comparison to, for example the Landweber algorithm, the algorithm additionally requires the pointwise evaluation of H_p and φ_p which can be done rather fast. Moreover, since the iteration procedure in Step 4 is just a convex combination, we can reuse $K(v^n - u^n)$ for the computation of Ku^{n+1}, so we have to compute only one evaluation of K and K^* in each iteration, respectively.

Remark 4.11. The term

$$D_n = \sum_{k=1}^{\infty} w_k \big(\varphi_p(u_k^n) - \varphi_p(v_k^n)\big) + \big(K^*(Ku^n - f)\big)_k (u_k^n - v_k^n)$$

in Step 3 of the algorithm can be used as an a-posteriori error bound on the distance to the minimizer, i.e. it holds $D_n \geq \tilde{\Psi}(u^n) - \min_{u \in \ell^2} \tilde{\Psi}(u)$, see Lemma 3.5. So one can use the stopping criterion $D_n < \varepsilon$ to assure that the minimal value is reached up to a certain tolerance $\varepsilon > 0$ in case of convergence, see the Appendix for details.

Remark 4.12. Note that if $p = 1$ the penalty functional

$$\Phi(u) = \sum_{k=1}^{\infty} w_k |u_k|$$

is non-differentiable. A common workaround for the lack of differentiability was to regularize the modulus function by the differentiable function

$$\varphi_{1,\varepsilon}(t) = \sqrt{t^2 + \varepsilon^2}$$

with a small $\varepsilon > 0$ (see e.g. [25] where this way was introduced for regularizing the TV norm). This always introduced some deviation to the real solution and posed numerical difficulties for very small ε. Especially the desired property of sparseness of the solution is lost.

In the algorithm presented above, we do in some sense the opposite: We modify the modulus function for large values in order to make the generalized conditional gradient method applicable. In this case, the modification is outside of the domain relevant for the minimization problem (2.1) and the solutions obtained are the exact sparse solutions.

5. Numerical experiments. To illustrate the convergence behavior of the iterated hard shrinkage algorithm as stated in Algorithm 2.2, we performed numerical tests on three linear model problems and compared the results to the iterated soft shrinkage algorithm introduced in [15] and [22]. Our primary aim is to demonstrate the applicability of the new algorithm, we thus perform the experiments for problems well-known in image processing and inverse problems.

5.1. Deblurring. The first model problem we tested is an image deblurring (deconvolution) problem with known kernel and penalization with respect to the coefficients in a Haar wavelet basis. The problem was discretized to a rectangular grid of points, i.e. we consider $u = \{u_{ij}\}$ with $1 \leq i \leq N$ and $1 \leq j \leq M$ and pose the minimization problem

$$\min_u \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{\big((u * g)_{ij} - f_{ij}\big)^2}{2} + \sum_{k=1}^{NM} w_k |\langle u, \psi_k \rangle|^p$$

where ψ_k is the discrete two-dimensional wavelet basis spanning the space $\mathbf{R}^{NM}$ and $*$ denotes the usual discrete convolution with a kernel $g = \{g_{ij}\}$.

For our computations, we used an out-of-focus kernel g with radius $r = 6$ pixels which has been normalized to $\sum |g| = 0.99$ such that the associated operator's norm is strictly below 1. The original image of size 256×256 pixels with values in $[0, 1]$ was blurred with that kernel and in one case disturbed with Gaussian noise of variance η

u_{orig} f_1 f_2

FIG. 5.1. *The images used as measured data f for the deblurring problem. From left to right, the original image, the blurred image without noise and the noisy blurred image used for test 1 and 2 respectively, are depicted.*

(see Figure 5.1). Then, Algorithm 2.2 as well as the iterated soft shrinkage procedure was applied for a suitable number of iterations with a wavelet decomposition up to level 5 and the following parameters:

1. $p = 1$, $w_k = 0.00002$, $\eta = 0$
2. $p = 1.75$, $w_k = 0.005$, $\eta = 0.025$

For the comparison of these two methods, we plotted the functional values at each iteration step. The results for the two deblurring tests are depicted in Figure 5.2.

Note that if $1 < p < 2$, we observed that the hard shrinkage iteration usually performs faster than the soft shrinkage algorithm although we did not optimize for computational speed. The reason for this is that the iterated soft shrinkage algorithm demands the solution of

$$x + \bar{w}p\,\mathrm{sgn}(x)|x|^{p-1} = y$$

for all basis coefficients in each iteration step which has to be done by numerical approximation. Experiments show that it is necessary to compute the solution sufficiently precise since otherwise an increase of the functional value is possible. This is a relatively time-consuming task even if the quadratically convergent Newton method is used. Alternatively, one can use tables to look up the required values, but this still has to be done for a wide range of numbers and up to a high precision. In comparison, the iterated hard shrinkage method (Algorithm 2.2), only the evaluation of $|x|^{1/(p-1)}$ is necessary which can usually be done significantly faster and requires no special implementation.

5.2. Backwards heat conduction. The second model problem we considered was solving the backwards heat equation in one dimension with sparsity constraints in a point basis. In this experiment we investigated the role of p and its influence on the performance of the algorithm.

The continuous model reads as follows: Consider an initial condition $u^0 \in \mathrm{L}^2([0,1])$ and the one-dimensional heat equation with Dirichlet boundary conditions:

$$\begin{aligned} u_t &= u_{xx} \quad \text{for } (t,x) \in [0,T] \times [0,1] \\ u(0,x) &= u^0(x) \\ u(t,0) &= u(t,1) = 0. \end{aligned}$$

With K we denote the operator which maps the initial condition onto the solution of the above equation at time T. The problem of finding the initial condition u^0 from

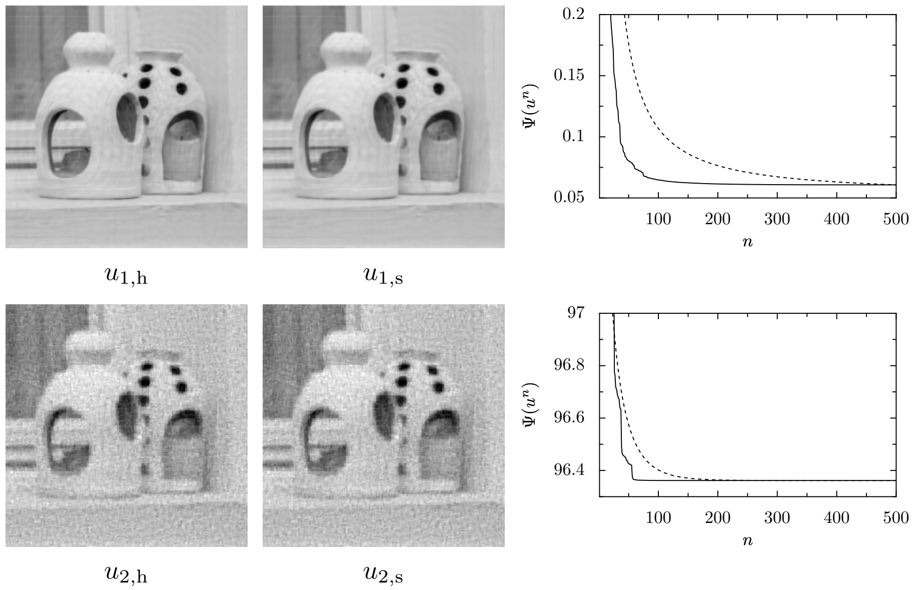

FIG. 5.2. *The results of the deblurring tests. In the left column you can see the reconstructed u after 500 iterations with the iterated hard shrinkage method while in the middle column, the reconstruction u after the same amount of iterations with the iterated soft shrinkage procedure is depicted. On the right hand side, a comparison of the descent of the functional values $\Psi(u^n)$ for both methods is shown. The solid and dashed lines represent the values of Ψ for the iterated hard and soft shrinkage procedure, respectively.*

the measurement of the heat distribution f at time T is thus formulated as solving

$$Ku^0 = f.$$

For the discretized model, we choose $u^0 = \{u_k\} \in \mathbf{R}^N$, data $f = \{f_j\} \in \mathbf{R}^M$ where u_k stands for the value of u^0 at point $x_k = (k-1)/(N-1)$. In the case of the heat equation, the solution matrix for the forward problem reads as

$$K_{kj} = \frac{2}{N} \sum_{l=1}^{\infty} e^{-\pi^2 l^2 T} \sin(\pi l x_k) \sin(\pi l x_j).$$

The minimization problem then reads as

$$\min_u \sum_{j=1}^{M} \frac{\big((Ku)_j - f_j\big)^2}{2} + \sum_{k=1}^{N} w_k |u_k|^p \ .$$

To test the algorithm, we created an initial distribution u^0 with one spike. The data $f = Ku + \delta$ is degraded with a relative error of 15% (see Figure 5.3).

We solved the above minimization problem with the iterated hard shrinkage algorithm for $w_k = 0.03$ and different values of p, namely

$$p_1 = 1,\ p_2 = 1.01,\ p_3 = 1.5.$$

As worked out in the Appendix the values D_n as defined in (3.5) can be used as a stopping criterion and in this experiment we stopped the iteration if D_n becomes

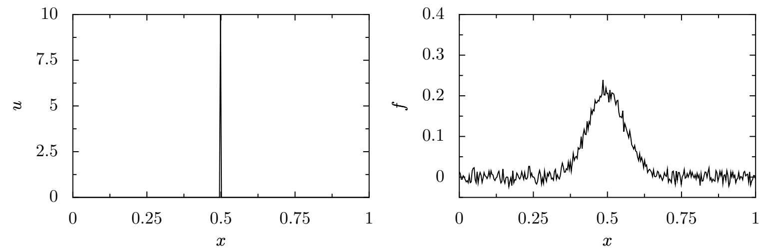

FIG. 5.3. *The data of the second model problem. From left to right: The spike initial heat distribution u and the heat distribution at time $T = .002$. (Note the different scaling.)*

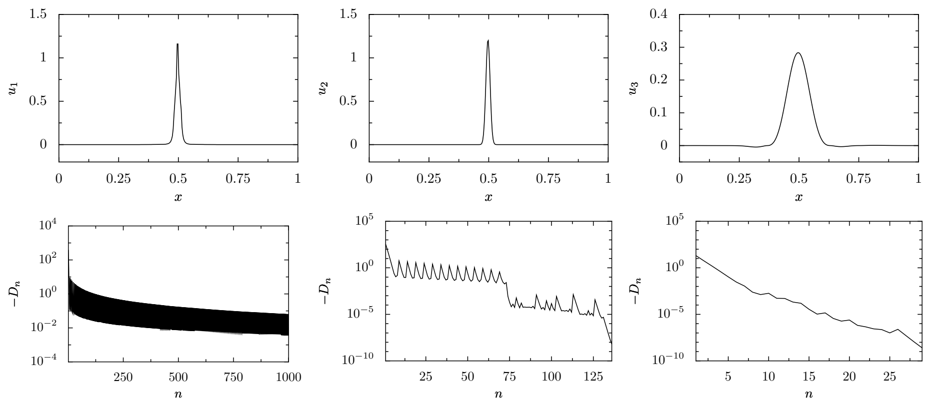

FIG. 5.4. *The results of the reconstruction of the initial condition. Top row from left to right: Reconstruction for $p = 1$, $p = 1.01$, and $p = 1.5$ respectively. Bottom row: The values of the estimator D_n in a logarithmic scale. (Note again different scaling.)*

smaller than 10^{-8} or the maximum number of 1000 iterations is reached. Figure 5.4 shows the results of the minimization process together with the estimators D_n. Note that the minimizers for $p = 1$ and $p = 1.01$ do not differ too much although the estimator D_n behaves very different: For $p = 1$ it oscillates heavily and is decaying slowly as the theory indicates. The slight change from $p = 1$ to $p = 1.01$ results in an estimator which is still oscillating but vanishing much faster and the algorithm stopped after 136 iterations. For $p = 1.5$ the sparsity of the reconstruction is lost but the algorithm terminated after just 29 iterations.

5.3. Inverse integration. In the last experiment we compared the iterative soft and hard shrinkage method in the case $p = 1$ with considerable sparsity of the solution, and with noisy data. The problem under consideration is inverse integration, i.e. the operator

$$Ku(t) = \int_0^t u(s)\mathrm{d}s, \quad s, t \in [0, 1], \quad u \in \mathrm{L}^2([0, 1]).$$

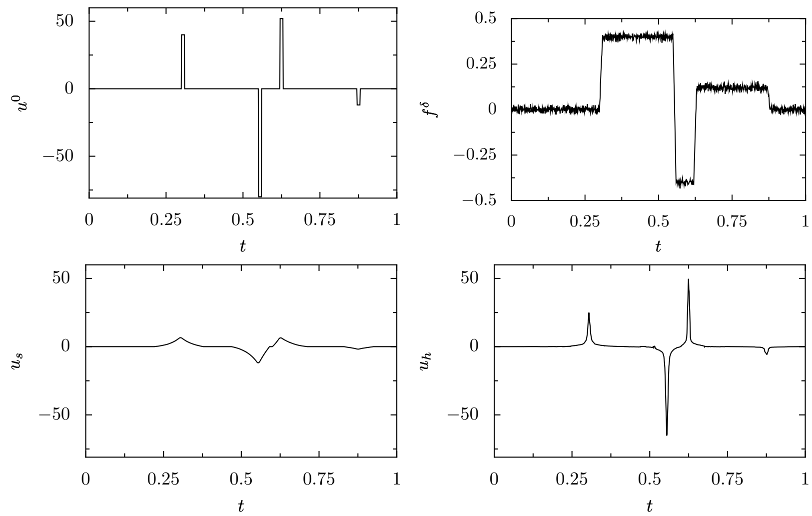

Fig. 5.5. *Top left: the true solution u^0, top right: the noisy data $f^\delta = Ku + \delta$. Bottom left: the result of the iterative soft shrinkage, bottom right: the result of the iterative hard shrinkage.*

The data f is given as $\{f(t_k)\}_{k=1,\dots,N}$ with $t_k = \frac{1}{N}k$. We discretized the operator K by the matrix

$$K = \frac{1}{N}\begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 1 & \dots & \dots & 1 \end{pmatrix}.$$

The true solution u^0 is given by small plateaus and hence the data $f^\delta = Ku^0 + \delta$ is a noisy function with steep linear ramps. We used $N = 1000$ and a relative error of 5%. The minimization problem reads as

$$\min_u \sum_{i=1}^{N} \frac{\big((Ku)_i - f_i^\delta\big)^2}{2} + \sum_{k=1}^{N} w_k |u_k|$$

which was solved with the iterative soft and hard shrinkage method for the parameter $w_k = .002$. The results after 1000 iterations are shown in Figure 5.5 while Figure 5.6 shows the decrease of the function Ψ for both methods.

6. Conclusion. We proposed a new algorithm for the minimization of functionals of type

$$\sum_k \frac{(Ku - f)_k^2}{2} + w_k |u_k|^p, \quad 1 \le p \le 2$$

in the infinite-dimensional setting. Our algorithm is based on iterated hard shrinkage. We established convergence rates for this algorithm, namely we proved convergence with rate $\mathcal{O}(n^{-1/2})$ for $p = 1$ and $\mathcal{O}(\lambda^n)$ for $1 < p \le 2$. We remark that the iterative hard shrinkage is a discontinuous algorithm, hence convergence rates are not at all easy to establish.

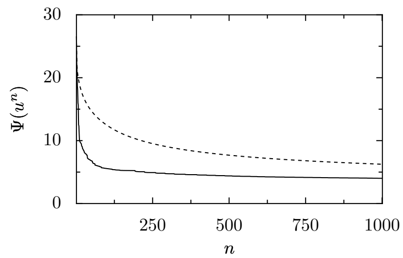

FIG. 5.6. *A comparison of the descent of the functional values $\Psi(u^n)$ for both methods is shown. The solid and dashed lines represent the values of Ψ for the iterated hard and soft shrinkage procedure, respectively.*

For finite-dimensional problems of the above type there are other algorithms with better performance (e.g. interior point methods, see [5, 3]), but none of them has a proper foundation for the infinite-dimensional setting. To our best knowledge the results stated here are the first results on convergence rates for a minimization algorithm of the above functional in infinite dimensions.

We emphasize that the convergence rate is only influenced by Φ and not by F, i.e. not by the operator K. For functionals $\Phi(u) = |u|_{H^s}$ one can expect similar convergence rates. Unfortunately, the case of total variation deblurring $\Phi(u) = |u|_{TV}$ seems not to fit into this context and further analysis in needed (while the case of the discrete TV-norm as treated in [4] goes well with this algorithm).

The change of the convergence rate from $p > 1$ to $p = 1$ is rather drastically: from convergence as λ^n to convergence as $n^{-1/2}$ – and this is observed in the numerical experiments. To provide an explanation: the contraction constant λ tends to 1 for $p \to 1$ which can been seen by analyzing the respective constants, especially $c(L)$, in the proof of Theorem 3.11. To speed up the minimization for $p = 1$ it could be of interest to use the minimizer of the minimization problem for p slightly larger than 1 as initial value u_0 for the iteration with $p = 1$. Another possibility is to decrease p during the iteration and use the framework of Γ-convergence to prove convergence of the algorithm.

The new iterative hard shrinkage algorithm works well in many cases since it seems to have ‘good ideas’ for finding ‘unconventional’ descent directions. On the other hand it sometimes runs into a situation where it can not find good ways for further descent (see Figure 5.2: some steps reduce the functional values drastically while sometimes the functional does not decrease much for many iterations). The iterative soft shrinkage, whereas, gives well descent in every step. Hence, a combination of both may share the good features of both.

As a side result we established an estimator D_n for the distance of the nth iterate to the minimizer which can be used as a stopping criterion for general iterative algorithms in case of convergence (see Remark A.3 in the Appendix).

7. Acknowledgement. We would like to thank both anonymous referees for their remarks and criticism – especially for shortening two proofs of the original manuscript. Kristian Bredies acknowledges support of the DFG under grant MA 1657/14-1.

Appendix. Convergence of the a-posteriori error bound. For numerical computations, it is often useful to have an estimate on some error so one can formulate stopping criteria for iterative algorithms. As already mentioned in Remark 4.11, the generalized conditional gradient algorithm (Algorithm 3.1) involves the estimates to the distance to the minimizer D_n according to (3.5). The following proposition shows that they also vanish in case of convergence and therefore $D_n < \varepsilon$ for some $\varepsilon > 0$ can be used as a stopping criterion for Algorithm 3.1. This stopping criterion can be formulated for general penalty functionals, therefore we prove it for the general setting.

PROPOSITION A.1. *Let Φ given according to Condition 3.2 and consider a sequence $\{u^n\}$ which is generated by Algorithm 3.1 for the solution of* (3.1).

Then

$$D_n = \Phi(u^n) - \Phi(v^n) + \langle K^*(Ku^n - f),\, u^n - v^n\rangle \to 0$$

for $n \to \infty$.

Proof. First observe that the descent property of Lemma 3.8 implies convergence of the functional values $\Psi(u^n)$, especially $\Psi(u^n) - \Psi(u^{n+1}) \to 0$. In Lemma 3.8, the estimate

$$\Psi(u^{n+1}) - \Psi(u^n) \leq \begin{cases} \frac{-D_n}{2} & \text{if } D_n \geq \|K(v^n - u^n)\|^2 \\ \frac{-D_n^2}{2\|K(v^n-u^n)\|^2} & \text{if } D_n \leq \|K(v^n - u^n)\|^2 \ . \end{cases}$$

which is slightly different to (3.8), is also proven. Remember that Condition 3.2 and the descent property imply $\|v^n - u^n\| \leq C$ (cf. the proof of Lemma 3.9), thus one can deduce

$$D_n \leq \max \Big\{ 2\big(\Psi(u^n) - \Psi(u^{n+1})\big), \sqrt{2}C\|K\|\big(\Psi(u^n) - \Psi(u^{n+1})\big)^{1/2} \Big\}$$

which proves the assertion since $D_n \geq 0$ and the right hand side obviously converges to 0 as $n \to \infty$. □

Remark A.2. If Φ fulfills Condition 3.2, then it is possible to compute $D(u)$ without the knowledge of v with the help of the conjugate functional Φ^* (see [13] for an introduction): The requirement that v is a solution of

$$\min_{v \in \mathcal{H}_1} \ \langle K^*(Ku - f),\, v\rangle + \Phi(v)$$

can equivalently be expressed by $-K^*(Ku - f) \in \partial\Phi(v)$ which is in turn, with the help of the Fenchel identity, equivalent to

$$-\Phi(v) - \langle K^*(Ku - f),\, v\rangle = \Phi^*\big(-K^*(Ku - f)\big) \ .$$

Plugging this into the definition of D in (3.5) gives

$$D(u) = \langle K^*(Ku - f),\, u\rangle + \Phi(u) + \Phi^*\big(-K^*(Ku - f)\big) \ , \tag{A.1}$$

a formula where v no longer appears.

Remark A.3. Equation (A.1) can also be used as a stopping criterion for general iterative minimization algorithms in case the algorithm converges to an optimal solution u^*:

Assume that $u^n \to u^*$ with $\Psi(u^n) \to \Psi(u^*) = \min_{u \in \mathcal{H}_1} \Psi(u)$ and Φ satisfies Condition 3.2. Then $\frac{1}{2}\|Ku^n - f\|^2 \to \frac{1}{2}\|Ku^* - f\|^2$ by continuity and consequently $\Phi(u^n) \to \Phi(u^*)$ by the minimization property. Now Condition 3.2 on Φ implies that

Φ^* is less than infinity on the whole space $\mathcal{H}_1$: Assume the opposite, i.e. the existence of a sequence $v^n \in \mathcal{H}_1$ as well as a $v^* \in \mathcal{H}_1$ such that

$$\sup_n \langle v^n, v^* \rangle - \Phi(v^n) = \infty \ .$$

This immediately implies that $v^* \neq 0$ and $\|v^n\| \to \infty$ since Φ has to be bounded from below (Φ is coercive). But moreover, Condition 3.2 implies $\Phi(v^n)/\|v^n\| \to \infty$ which means that for each $C > \|v^*\|$ one can find an n_0 such that $-\Phi(v^n) \leq -C\|v^n\|$ for $n \geq n_0$ and, consequently,

$$\sup_{n \geq n_0} \langle v^n, v^* \rangle - \Phi(v^n) \leq (\|v^*\| - C)\|v^n\| \leq 0$$

which is a contradiction.

It follows that Φ^* is less than infinity everywhere and hence continuous (see e.g. [20]). So additionally, $\Phi^*\bigl(-K^*(Ku^n - f)\bigr) \to \Phi^*\bigl(-K^*(Ku^* - f)\bigr)$ and the continuity of the scalar product leads to

$$\lim_{n\to\infty} D_n = \langle K^*(Ku^* - f), u^* \rangle + \Phi(u^*) + \Phi^*\bigl(-K^*(Ku^* - f)\bigr) = 0 \ .$$

Remark A.4. Let us show how the error bound D can be computed in the modified problem (4.1).

We first state the conjugate functionals Φ^* associated with the penalty functionals Φ. With some calculation one obtains for the conjugates of φ_p as defined in (2.2) that

$$\varphi_p^*(x^*) = \begin{cases} (p-1)\bigl|\frac{x^*}{p}\bigr|^{p'} & \text{if } |x^*| \leq pS_0^{p-1} \\ \frac{S_0^{2-p}}{2p}(x^*)^2 + \bigl(\frac{p}{2} - 1\bigr)S_0^p & \text{if } |x^*| > pS_0^{p-1} \end{cases}$$

where p' denotes the dual exponent, i.e. $\frac{1}{p} + \frac{1}{p'} = 1$ and we formally define that $\frac{1}{\infty} = 0$.

Since the functional Φ is defined by pointwise summation in ℓ^2, we can also take the conjugate pointwise. This gives

$$\Phi^*(u^*) = \sum_{k=1}^{\infty} w_k \varphi_p^*\Bigl(\frac{u_k^*}{w_k}\Bigr)$$

with the φ_p^* as above. Eventually, the error bound at some $u \in \ell^2$ can be expressed by

$$D(u) = \sum_{k=1}^{\infty} \bigl(K^*(Ku - f)\bigr)_k u_k + w_k\varphi_p(u_k) + w_k\varphi_p^*\Bigl(\frac{-\bigl(K^*(Ku - f)\bigr)_k}{w_k}\Bigr) \ .$$

REFERENCES

[1] Julien Bect, Laure Blanc-Féraud, Gilles Aubert, and Antoine Chambolle, *A l^1-unified variational framework for image reconstruction*, in Proceeding of the European Conference on Computer Vision, Tomás Pajdla and Jiri Matas, eds., vol. 3024 of Lecture Notes in Computer Science, Springer, 2004, pp. 1–13.

[2] Kristian Bredies, Dirk A. Lorenz, and Peter Maass, *A generalized conditional gradient method and its connection to an iterative shrinkage method*. Accepted for publication in *Computational Optimization and Applications*, 2005.

[3] Emmanuel J. Candès and Terence Tao, *Decoding by linear programming*, IEEE Transaction on Information Theory, 51 (2004), pp. 4203–4215.
[4] Antonin Chambolle, *An algorithm for total variation minimization and applications*, Journal of Mathematical Imaging and Vision, 20 (2004), pp. 89–97.
[5] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders, *Atomic decomposition by basis pursuit*, SIAM Journal on Scientific Computing, 20 (1998), pp. 33–61.
[6] Patrick L. Combettes and Valérie R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., 4 (2005), pp. 1168–1200.
[7] Ingrid Daubechies, *Ten Lectures on Wavelets*, vol. 61 of CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM (Society for Industrial and Applied Mathematics), 1992.
[8] Ingrid Daubechies, Michel Defrise, and Christine De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Communications in Pure and Applied Mathematics, 57 (2004), pp. 1413–1457.
[9] David L. Donoho and Michael Elad, *Optimally-sparse representation in general (non-orthogonal) dictionaries via ℓ^1 minimization*, Proceedings of the National Academy of Sciences, 100 (2003), pp. 2197–2202.
[10] David L. Donoho, Michael Elad, and Vladimir Temlyakov, *Stable recovery of sparse overcomplete representations in the presence of noise*, IEEE Transactions on Information Theory, 52 (2006), pp. 6–18.
[11] Joseph C. Dunn, *Rates of convergence for conditional gradient algorithms near singular and nonsingular extremals*, SIAM Journal on Control and Optimization, 17 (1979), pp. 187–211.
[12] ———, *Convergence rates for conditional gradient sequences generated by implicit step length rules*, SIAM Journal on Control and Optimization, 18 (1980), pp. 473–487.
[13] Ivar Ekeland and Roger Temam, *Convex Analysis and Variational Problems*, North-Holland, Amsterdam, 1976.
[14] Michael Elad, Boaz Matalon, and Michael Zibulevsky, *Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization*. To appear in *Journal on Applied and Computational Harmonic Analysis*.
[15] Mário A. T. Figueiredo and Robert D. Nowak, *An EM algorithm for wavelet-based image restoration*, IEEE Transactions on Image Processing, 12 (2003), pp. 906–916.
[16] Jean-Jacques Fuchs, *Recovery of exact sparse representations in the presence of bounded noise*, IEEE Transactions on Information Theory, 51 (2005), pp. 3601–3608.
[17] Jean-Jacques Fuchs and Bernard Delyon, *Minimal L_1-norm reconstruction function for oversampled signals: Applications to time-delay estimation.*, IEEE Transactions on Information Theory, 46 (2000), pp. 1666–1673.
[18] Rémi Gribonval and Morten Nielsen, *Sparse representation in unions of bases*, IEEE Transactions of Information Theory, 49 (2003), pp. 3320–3325.
[19] Bhaskar D. Rao, *Signal processing with the sparseness constraint*, in Proceedings of International Conference on Acoustics, Speech and Signal Processing, vol. 1, Seattle, Washington, May 1998, pp. 369–372.
[20] Ralph Edwin Showalter, *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*, vol. 49 of Mathematical Surveys and Monographs, American Mathematical Society, 1997.
[21] Jean-Luc Starck, Emmanuel J. Candès, and David L. Donoho, *Astronomical image representation by the curvelet transform*, Astronomy and Astrophysics, 398 (2003), pp. 785–800.
[22] Jean-Luc Starck, Mai K. Nguyen, and Fionn Murtagh, *Wavelets and curvelets for image deconvolution: a combined approach*, Signal Processing, 83 (2003), pp. 2279–2283.
[23] Gerd Teschke, *Multi-frames in thresholding iterations for nonlinear operator equations with mixed sparsity constraints*, tech. report, Preprint series of the DFG priority program 1114 "Mathematical methods for time series analysis and digital image processing", July 2005.
[24] Joel A. Tropp, *Just relax: Convex programming methods for identifying sparse signals in noise*, IEEE Transactions on Information Theory, 52 (2006), pp. 1030–1051.
[25] Curtis R. Vogel and Mary E. Oman, *Iterative methods for total variation denoising*, SIAM Journal on Scientific Computing, 17 (1996), pp. 227–238.