Regularization of linear ill-posed problems by the augmented Lagrangian method and variational inequalities

K Frick and M Grasmair

1 Institute for Mathematical Stochastics, University of Göttingen, Goldschmidtstraße 7, 37077 Göttingen, Germany
2 Computational Science Center, University of Vienna, Nordbergstraße 15, 1090 Vienna, Austria

E-mail: frick@math.uni-goettingen.de and markus.grasmair@univie.ac.at

Received 3 April 2012, in final form 26 June 2012
Published 1 October 2012
Online at stacks.iop.org/IP/28/104005

Abstract

We study the application of the augmented Lagrangian method to the solution of linear ill-posed problems. Previously, linear convergence rates with respect to the Bregman distance have been derived under the classical assumption of a standard source condition. Using the method of variational inequalities, we extend these results in this paper to convergence rates of lower order, both for the case of an a priori parameter choice and an a posteriori choice based on Morozov’s discrepancy principle. In addition, our approach allows the derivation of convergence rates with respect to distance measures different from the Bregman distance. As a particular application, we consider sparsity promoting regularization, where we derive a range of convergence rates with respect to the norm under the assumption of restricted injectivity in conjunction with generalized source conditions of Hölder type.

1. Introduction

We aim for the solution of the problem

$$\inf_{u \in X} J(u) \quad \text{s.t.} \quad Ku = g,$$

(1)

where $K : X \to H$ is a linear and bounded mapping between a Banach space $X$ and a Hilbert space $H$ and where $J : X \to \mathbb{R}$ is convex and lower semi-continuous. We are particularly interested in the case when the right-hand side in the linear constraint is not at hand but only an approximation $g^\delta$, such that

$$\|g - g^\delta\| \leq \delta$$

(2)

for some $\delta > 0$. A possible method for computing a stable approximation of solutions of (1) is the augmented Lagrangian method (ALM), an iterative method that, for a given initial value...
Inverse Problems 28 (2012) 104005  K Frick and M Grasmair

$p^k_0 \in H$ and for $k = 1, 2, \ldots$, computes

$$u^k_\delta \in \arg\min_{u \in X} \frac{\tau_k}{2} \| Ku - g^\delta \|^2 + J(u) - \langle p^k_{k-1}, Ku - g^\delta \rangle.$$  \hspace{1cm} (3a)

$$p^k_{k-1} = p^k_{k-1} + \tau_k (g^\delta - Ku^k_{\delta}).$$  \hspace{1cm} (3b)

Here, $\{\tau_k\}_{k \in \mathbb{N}}$ denotes a pre-defined sequence of positive parameters, such that

$$t_n := \sum_{k=1}^{n} \tau_k \to \infty \quad \text{as} \quad n \to \infty.$$  

The ALM was originally introduced by Hestenes (1969) and Powell (1969) (under the name method of multipliers) as a solution method for problems of type (1) in the Euclidean space with exact right-hand side $g$. Since then the ALM was developed further in various directions; see e.g. Fortin and Glowinski (1983), Ito and Kunisch (2008) and the references therein. In particular, the method has been generalized to Hilbert and also Banach spaces (note that the infinite-dimensional setting has already been briefly discussed by Hestenes (1969)).

In the context of inverse problems, the ALM was first considered for the special case when $X$ is a Hilbert space and $J$ is a quadratic functional, i.e. $J(u) = \frac{1}{2} \| Lu \|^2$ for a densely defined and closed linear operator $L : D(L) \subset X \to \tilde{H}$, where $\tilde{H}$ is some further Hilbert space (here we set $J(u) = +\infty$ if $u \notin D(L)$). For this special case, it is readily seen that the ALM can be rewritten as

$$u^k_\delta = \arg\min_{u \in X} \left[ \tau_k \| Ku - g^\delta \|^2 + \| L(u - u^k_{\delta-1}) \|_{\tilde{H}}^2 \right].$$  \hspace{1cm} (4)

The analysis of iteration (4) dates back to the papers by Krasnosel’skii (1960) and Krzanev (1973). The case when $L \equiv \text{Id}$ is referred to as the iterated Tikhonov method and has been studied by Lardy (1975), Brill and Schock (1987), Hanke and Groetsch (1998) and Engl et al (1996). The regularization scheme that results for $K \equiv \text{Id}$ is termed iterated Tikhonov–Morozov method and amounts to stably evaluate the (possibly unbounded) operator $L$ at $g$ given only an approximation $g^\delta$ that satisfies (2). For detailed analysis, see e.g. Groetsch and Scherzer (2000) and Groetsch (2007).

A generalization of the iteration in (4) for total-variation-based image reconstruction has been established by Osher et al (2005) under the name Bregman iteration and convergence properties were studied by Burger et al (2007). Frick and Scherzer (2010) pointed out that the Bregman iteration and the iterated Tikhonov–Morozov method are the special instances of the ALM as it is stated in (3) and developed an improved convergence analysis. Frick et al (2011) also studied Morozov’s discrepancy principle (Morozov 1967) for the ALM. The application of the ALM for the regularization of nonlinear operators has been considered by Bachmayr and Burger (2009) and Jung et al (2011).

Up to now, convergence rates for the ALM (in the context of inverse problems) have only been derived under the assumption that the solutions $u^\dagger$ of (1) satisfy the standard source condition (Burger and Osher 2004)

$$K^* p^\dagger \in \partial J(u^\dagger) \quad \text{for some} \quad p^\dagger \in H.$$  \hspace{1cm} (5)

Here, $K^* : H \to X^*$ denotes the adjoint operator of $K$ and $\partial J(u^\dagger)$ is the subdifferential of $J$ at $u^\dagger$. This typically results in a convergence rate of $\delta$ with respect to the Bregman distance (for a definition of the subdifferential and the Bregman distance, see section 2). In this paper, we extend these results to convergence rates of lower order by replacing (5) by variational inequalities. The analysis will apply for both a priori and a posteriori parameter-selection rules, where the latter will be realized by Morozov’s discrepancy principle. In addition, our
approach allows the derivation of convergence rates with respect to distance measures different from the Bregman distance.

The rest of the paper is organized as follows. In section 2, we state basic assumptions and review tools from the convex analysis that are essential for our analysis. In section 3, we establish variational inequalities and prove that these are sufficient for lower order convergence rates for the ALM with suitable a priori stopping rules. In section 4, we prove the same convergence rates when Morozov’s discrepancy principle is employed as an a posteriori stopping rule. In section 5, we finally consider some examples that clarify the connection of the variational inequalities discussed in section 3 and more classic notions of source conditions, such as the standard source condition (5) or Hölder-type conditions. Moreover, we show for the particular scenario of sparsity promoting regularization how our approach can be used to derive convergence rates with respect to the norm.

2. Assumptions and mathematical prerequisites

In this section, we fix some basic assumptions and review basic notions and facts from convex analysis. We will henceforth denote by $\partial J(u_0)$ the subdifferential of $J$ at $u_0 \in X$, i.e. the set of all $\xi \in X^*$, such that $J(u) \geq J(u_0) + \langle \xi, u - u_0 \rangle_{X^*,X}$, for all $u \in X$.

In this case, we call $\xi$ a subgradient of $J$ at $u_0$. We denote by $K^* : H \to X^*$ the adjoint operator of $K$, where we identify the Hilbert space $H$ with its dual $H^*$ by means of Riesz’ representation theorem. Assumption 2.1 guarantees that solutions of (1) exist for all $g \in D(J)$ and that the iteration (3) is well defined. The proof is analogous to Frick and Scherzer (2010, lemma 3.1); the only difference is that there also the space $X$ is assumed to be a Hilbert space.

Recall that the Legendre–Fenchel conjugate $J^* : X^* \to \mathbb{R}$ of $J$ is defined by $J^*(x^*) = \sup_{x \in X} \langle x^*, x \rangle_{X^*,X} - J(x)$. The dual problem to (1) is then defined by (see Aubin (1979, section 5.2.9))

$$\inf_{p \in H} [J^*(K^*p) - \langle p, g \rangle].$$

Sufficient and necessary conditions for guaranteeing the existence of a (constrained) minimizer $u^i \in X$ of (1) and a minimizer $p^i \in H$ of (6) are the Karush–Kuhn–Tucker conditions (see Aubin (1979, sections 5.2.4 and 5.2.6)), which read

$$K^*p^i \in \partial J(u^i) \quad \text{and} \quad Ku^i = g.$$
From an inverse problems perspective, these conditions are understood as source conditions (Burger and Osher 2004) that delimit a class of particular regular solutions \( u^j \) of (1) that can be reconstructed from noisy data at a certain rate depending on the noise level \( \delta \). If the source condition (7) does not hold, then solutions of (1) may still exist (e.g. if assumption 2.1 holds), whereas (6) has no solutions. The value of (6), though, will still be finite.

**Lemma 2.2.** Suppose that assumption 2.1 holds and let \( u^j \in X \) be a solution of (1). Then,

\[
\inf_{p \in H} [J^*(K^* p) - \langle p, g \rangle] = -J(u^\dagger).
\]

**Proof.** Define a function \( \Gamma : X \times H \to \bar{R} \) by setting \( \Gamma(u, p) = J(u) \) if \( Ku = g + p \) and \( G(u, p) = +\infty \) otherwise. According to Ekeland and Temam (1976, chapter III, proposition 2.1) the assertion holds, if the function \( p \mapsto h(p) = \inf_{u \in X} \Gamma(u, p) \) is finite and lower semicontinuous at \( p = 0 \). Since \( p(0) = J(u^j) < \infty \), it remains to prove lower semicontinuity. Therefore, let \( \{p_k\}_{k \in N} \) be a sequence in \( H \), such that \( p_k \to 0 \). Without loss of generality, we may, after possibly passing to a subsequence, assume that \( h(p_k) < \infty \) for every \( k \), which amounts to saying that the equation \( Ku = g + p_k \) has a solution \( u_k \in X \) satisfying \( J(u_k) < \infty \). In addition, because of assumption 2.1, we can choose \( u_k \) such that the infimum in the definition of \( h \) is realized at \( u_k \), i.e. \( h(p_k) = \Gamma(u_k, p_k) \).

Now, if \( J(u_k) \to \infty \) as \( k \to \infty \), nothing remains to be proven. Thus, we can assume that there exists a subsequence of \( \{u_k\} \), such that \( \sup_{k \in \mathbb{N}} J(u_k) < \infty \). It is not restrictive to assume that \( \lim_{k \to \infty} J(u_k) = \lim \inf_{k \to \infty} J(u_k) \). Moreover, we observe that \( \|K u_k - g\|^2 = \|p_k\|^2 \) is bounded, since \( p_k \to 0 \). Thus, it follows from assumption 2.1 that there exists a further subsequence \( \{u_{k'}\} \), such that \( u_{k'} \to \hat{u} \) for some \( \hat{u} \in X \). This implies that \( Ku_{k'} \to K\hat{u} = g \), and the lower semicontinuity and convexity of \( J \) finally proves that

\[
\lim \inf_{k \to \infty} h(p_k) = \lim_{k' \to \infty} J(u_{k'}) = \lim \inf_{k' \to \infty} J(\hat{u}) \geq J(\hat{u}) = J(u^\dagger) = h(0). \quad \square
\]

A relation similar to the duality relation between the optimization problems (1) and (6) can be established for the ALM: as first observed by Rockafellar (1974), the dual sequence \( \{p^0_0, p^1_0, \ldots\} \) generated by the ALM can be characterized by the proximal point method (PPM).

To be more precise, for all \( k \geq 1 \),

\[
p^k_k = \arg \min_{p \in H} \left[\frac{1}{2} \left\| p - p^k_{k-1}\right\|^2 + \tau_k (J^*(K^* p) - \langle p, g^k \rangle)\right].
\]

The PPM was introduced by Martinet (1970) for minimizing a convex functional, which in the present situation is the dual functional (6). The sequence \( \{p^k_k\} \) generated by the PPM is known to converge weakly to a solution of (6) if it exists, i.e. when (7) holds. If this is not the case, then still \( J^*(K^* p^k_k) - \langle p^k_k, g^k \rangle \) converges to the value of the program (6) which, in the general case, may be \(-\infty\), of course.

### 3. Convergence rates

The classical analysis of the ALM within the context of optimization assumes that the right-hand side of the equation \( Ku = g \) is given exactly, i.e. \( \delta = 0 \). Under this assumption, the iterates of the ALM converge to the \( J \)-minimizing solution of \( Ku = g \) as \( n \to \infty \), provided there exists any solution \( u \) of this equation satisfying \( J(u) < \infty \); see, e.g., the results by Fortin and Glowinski (1983) and Ito and Kunisch (2008).

Within the context of inverse problems, however, the right-hand side is not known exactly but only approximately with some known error bound \( \delta > 0 \). That is, we are given \( g^\delta \in H \)
with $\|g^n - g\| \leq \delta$. Still, one wants to find an approximation of the solution $u^\dagger$ of the true equation $Ku = g$. In the case where the operator equation is ill-posed this is only possible, if the iteration is stopped well before the iterates converge. Moreover, the stopping index of the iteration has to depend on the noise level $\delta$. In this setting, one can prove the following convergence result.

**Theorem 3.1.** Assume that the equation $Ku = g$ has a $J$-minimizing solution $u^\dagger \in D(J)$. If $n = n(\delta)$ is chosen in such a way that

$$
\lim_{\delta \to 0} \delta^2 n(\delta) = 0 \quad \text{and} \quad \lim_{\delta \to 0} n(\delta) = +\infty.
$$

Then, $\lim_{\delta \to 0} \|Ku_{n(\delta)} - g\| = 0$ and $\lim_{\delta \to 0} J(u_{n(\delta)}^\dagger) = J(u^\dagger)$. In particular, every weak cluster point of the weakly compact set $\{u_{n(\delta)}^\dagger\}_{\delta > 0}$ is a $J$-minimizing solution of $Ku = g$.

**Proof.** See Frick and Scherzer (2010, theorem 5.3). \hfill \Box

The previous result does not include any estimate of the speed of the convergence of the approximate solutions $u_{n(\delta)}^\dagger$ to $J$-minimizing solutions of $Ku = g$. In order to obtain such an estimate, it is necessary to impose certain regularity conditions on the true solution of the equation. It is well known that linear convergence rates (with respect to the Bregman distance) for iterates of the ALM can be proven if the source condition (7) holds (cf Burger et al. (2007) and Frick and Scherzer (2010)). In this section, we prove lower order rates of convergence in the case, when the source condition (7) does not hold. Instead, we impose weaker regularity conditions on solutions $u^\dagger$ of (1) in terms of variational inequalities. We formulate this in the following.

**Assumption 3.2.** We are given an index function $\Phi : [0, \infty) \to [0, \infty)$, i.e. a non-negative continuous function that is strictly increasing and concave with $\Phi(0) = 0$. Moreover, $D : X \times X \to [0, \infty]$ satisfies $D(u, u) = 0$ whenever $u \in X$, and $u^\dagger$ is a solution of (1) such that

$$
D(u, u^\dagger) \leq J(u) - J(u^\dagger) + \Phi(\|Ku - g\|^2) \quad \text{for all } u \in X.
$$

We denote by $\Psi$ the Legendre–Fenchel conjugate of $\Phi^{-1}$.

A typical choice is $D(u, v) = \beta D_J^\dagger(u, v)$, where $\beta \in (0, 1]$ and

$$
D_J^\dagger(u, v) = J(u) - J(v) - \langle \xi, u - v \rangle_{X^*, X}
$$

is the Bregman distance of $u$ and $v$ w.r.t. $\xi \in \partial J(v)$. With this, (9) is equivalent to the condition

$$
\langle \xi, u^\dagger - u \rangle_{X^*, X} \leq (1 - \beta) D_J^\dagger(u, u^\dagger) + \Phi(\|Ku - g\|^2)
$$

for all $u \in X$. In this form, variational inequalities have been introduced by Hofmann et al. (2007) and Scherzer et al. (2009) with $\Phi(s) = \sqrt{s}$, and for general index functions by Boţ and Hofmann (2010) and Grasmair (2010).

The following theorem asserts that condition (9) in assumption 3.2 imposes sufficient smoothness on the true solution $u^\dagger$ that the iterates of the ALM approach $u^\dagger$ with a certain rate (that depends on $\Phi$).

**Theorem 3.3.** Let assumptions 2.1 and 3.2 hold. Then, there exists a constant $C > 0$, such that

$$
D(u^\delta, u^\dagger) \leq C \delta^2 \left( \Phi \left( \frac{16}{t_n} \right) + \delta^2 \right)
$$

(12)
and
\[ \|Ku_n^\delta - g\|^2 \leq C \left( \psi \left( \frac{16}{t_n} \right) + \delta^2 \right). \]  
(13)

In particular, if \( t_n \gg \frac{1}{\psi^{-1}(\delta)} \), then
\[ D(u_n^\delta, u^\delta) = O \left( \frac{\delta^2}{\psi^{-1}(\delta)} \right) \quad \text{and} \quad \|Ku_n^\delta - g\|^2 = O(\delta^2). \]  
(14)

**Proof.** The theorem is a consequence of the two lemmas mentioned below. Combining the estimates derived in lemmas 3.4 and 3.5, it follows that
\[ D(u_n^\delta, u^\delta) \leq \tilde{C} t_n \left( \psi \left( \frac{16}{t_n} \right) + \delta^2 + \frac{1}{2} \psi \left( \frac{2}{t_n} \right) \right) \]
for some \( \tilde{C} > 0 \) and a similar estimate holds for \( \|Ku_n^\delta - g\|^2 \). Now note that the fact that \( \Phi \) is an index function implies that \( \Psi = (\Phi^{-1})^* \) is convex, non-negative and \( \Psi(0) = 0 \). As a consequence, \( \Psi \) is non-decreasing and sub-additive on \([0, +\infty)\). Thus, inequalities (12) and (13) follow with \( C = \tilde{C} + 1/2 \). Finally, estimate (14) follows from the sub-additivity of \( \Psi \).

**Lemma 3.4.** Let assumptions 2.1 and 3.2 hold and define for \( p \in H, t > 0 \) and \( \delta \geq 0 \):
\[ \psi(p, t, \delta) = \left( t\psi(16/t) + t\delta^2 + J^*(K^*p) + J(u^\delta) - \langle p, g \rangle + \frac{\|p\|^2}{2t} \right). \]
Then, there exists a constant \( C > 0 \), such that
\[ D(u_n^\delta, u^\delta) \leq C\psi(p, t_n, \delta) \quad \text{and} \quad \|Ku_n^\delta - g\|^2 \leq C \frac{\psi(p, t_n, \delta)}{t_n} \]
(15)
for all \( p \in H \).

**Proof.** Without loss of generality, we assume that \( p^\delta_0 = 0 \) and we shall use the abbreviation \( G(p, g) = J^*(K^*p) - \langle p, g \rangle \). In Güler (1991, lemma 2.1), it was proved that for all \( p \in V \)
\[ \frac{t_n}{2} \|p_n^\delta - p_{n-1}^\delta\|^2 \leq G(p, g^\delta) - G(p_n^\delta, g^\delta) - \frac{\|p - p_n^\delta\|^2}{2t_n} + \frac{\|p\|^2}{2t_n}. \]
(16)

Since \( G(p, g^\delta) - G(p_n^\delta, g^\delta) = G(p, g) - G(p_n^\delta, g) + \langle p - p_n^\delta, g - g^\delta \rangle \) and \( p_n^\delta - p_{n-1}^\delta = \tau_n(g^\delta - Ku_n^\delta) \), this implies that
\[ \frac{t_n}{2} \|Ku_n^\delta - g\|^2 \leq G(p, g) - G(p_n^\delta, g) - \frac{\|p - p_n^\delta\|^2}{2t_n} + \frac{\|p\|^2}{2t_n} + \langle p - p_n^\delta, g - g^\delta \rangle \]
\[ \leq G(p, g) + J(u^\delta) - \frac{\|p - p_n^\delta\|^2}{2t_n} + \frac{\|p\|^2}{2t_n} + \langle p - p_n^\delta, g - g^\delta \rangle, \]
(17)
where the second inequality follows from lemma 2.2. Setting \( p = p_n^\delta \), this proves that
\[ \frac{t_n}{2} \|Ku_n^\delta - g\|^2 \leq J^*(K^*p_n^\delta) - \langle p_n^\delta, g \rangle + J(u^\delta) + \frac{\|p_n^\delta\|^2}{2t_n}. \]
Since \( K^*p_n^\delta \in \partial J(u_n^\delta) \), we observe that \( J^*(K^*p_n^\delta) + J(u_n^\delta) = \langle K^*p_n^\delta, u_n^\delta \rangle \) and conclude that
\[ \frac{t_n}{2} \|Ku_n^\delta - g\|^2 \leq J(u^\delta) - J(u_n^\delta) + \langle p_n^\delta, Ku_n^\delta - g \rangle + \frac{\|p_n^\delta\|^2}{2t_n} \]
\[ = J(u^\delta) - J(u_n^\delta) + \langle p_n^\delta, Ku_n^\delta - g \rangle + \langle p_n^\delta, g^\delta - g \rangle + \frac{\|p_n^\delta\|^2}{2t_n}. \]
Applying Young's inequality \((a, b) \leq \|a\|^2/2 + \|b\|^2/2\) first with \(a = \sqrt{2/t_n}p_n^\delta\) and \(b = (Ku_n^\delta - g^\delta)/\sqrt{t_n/2}\), and then with \(a = p_n^\delta/\sqrt{t_n}\) and \(b = \sqrt{t_n}(g^\delta - g)\), we obtain
\[
\frac{t_n}{4} \|Ku_n^\delta - g^\delta\|^2 \leq J(u^\dagger) - J(u_n^\dagger) + \|p_n^\delta, g^\delta - g\| + \frac{3}{2} \|p_n^\delta\|^2/2t_n.
\]
\[
\leq J(u^\dagger) - J(u_n^\dagger) + \frac{\delta^2 t_n}{2} + \frac{2 \|p_n^\delta\|^2}{t_n}.
\]
Summarizing, we find that
\[
\|Ku_n^\delta - g^\delta\|^2 \leq \frac{4}{t_n} (J(u^\dagger) - J(u_n^\dagger)) + 2\delta^2 + \frac{8 \|p_n^\delta\|^2}{t_n^2}.
\]
Now, we observe from (9) that \(J(u^\dagger) - J(u_n^\dagger) \leq -D(u_n^\delta, u^\dagger) + \Phi(\|Ku_n^\delta - g\|^2)\). Plugging this inequality into the above estimate yields
\[
\|Ku_n^\delta - g^\delta\|^2 + \frac{4}{t_n} D(u_n^\delta, u^\dagger) \leq \frac{4}{t_n} \Phi(\|Ku_n^\delta - g\|^2) + 2\delta^2 + \frac{8 \|p_n^\delta\|^2}{t_n^2}.
\]
Since \(\Psi\) is the Legendre–Fenchel conjugate of \(t \mapsto \Phi^{-1}(t)\), i.e. \(\Psi(s) = \sup_{t \geq 0} st - \Phi^{-1}(t)\), it follows that \(st \leq \Psi(s) + \Phi^{-1}(t)\) for all \(s, t \geq 0\), and in particular, for \(t = \Phi(r)\), that \(s\Phi(r) \leq \Psi(s) + r\) for all \(s, r \geq 0\). Setting \(s = 16/t_n\) and \(r = \|Ku_n^\delta - g^\delta\|^2\) gives
\[
\frac{4}{t_n} \Phi(\|Ku_n^\delta - g\|^2) = \frac{1}{4} \frac{16}{t_n} \Phi(\|Ku_n^\delta - g\|^2)
\leq \frac{1}{4} \Psi \left( \frac{16}{t_n} \right) + \frac{1}{4} \|Ku_n^\delta - g\|^2
\leq \frac{1}{4} \Psi \left( \frac{16}{t_n} \right) + \frac{1}{2} \|Ku_n^\delta - g\|^2 + \frac{\delta^2}{2}.
\]
Combining this with (18) yields
\[
\frac{1}{2} \|Ku_n^\delta - g^\delta\|^2 + \frac{4}{t_n} D(u_n^\delta, u^\dagger) \leq \frac{1}{4} \Psi \left( \frac{16}{t_n} \right) + \frac{5\delta^2}{2} + \frac{8 \|p_n^\delta\|^2}{t_n^2}.
\]
Finally, we observe again from (16) that for all \(p \in H\)
\[
\|p - p_n^\delta\|^2 \leq \frac{G(p, g^\delta) - G(p_n^\delta, g^\delta)}{t_n} + \frac{\|p\|^2}{2t_n^2}
\leq \frac{G(p, g) - G(p_n^\delta, g)}{t_n} + \frac{1}{t_n} \|p - p_n^\delta, g - g^\delta\| + \frac{\|p\|^2}{2t_n^2}
\leq \frac{G(p, g) - \inf_{q \in H} G(q, g)}{t_n} + \frac{\|p - p_n^\delta\|^2}{4t_n^2} + \frac{\|p\|^2}{4t_n^2} + \delta^2 + \frac{\|p\|^2}{4t_n^2}.
\]
This shows that
\[
\|p_n^\delta\|^2 \leq \frac{\|p - p_n^\delta\|^2}{8t_n^2} + \frac{\|p\|^2}{4t_n^2}
\leq \frac{G(p, g) - \inf_{q \in H} G(q, g)}{t_n} + \delta^2 + \frac{3 \|p\|^2}{4t_n^2}.
\]
Combining (20) with (19) and applying lemma 2.2 finally gives

\[
\frac{1}{2} \| Ku_n^\delta - g \|^2 + \frac{4}{t_n^2} D(u_n^\delta, u^\dagger) \leq \frac{1}{4} \psi \left( \frac{16}{t_n} \right) + \frac{5\delta^2}{2} + \frac{8 \| p_n^\delta \|^2}{t_n^2} \\
\leq \frac{1}{4} \psi \left( \frac{16}{t_n} \right) + 64 \frac{G(p, g) - \inf_{u \in V} G(q, g)}{t_n} + \frac{13\delta^2}{2} + \frac{48 \| p \|^2}{t_n^2}.
\]

Lemma 3.5. Let assumptions 2.1 and 3.2 hold. Then,

\[
\inf_{p \in H} \left[ J^*(K^* p) + J(u^\dagger) - \langle p, g \rangle + \frac{\| p \|^2}{2t} \right] \leq \frac{t}{2} \psi \left( \frac{2}{t} \right).
\]

Proof. Classical duality theory (see Ekeland and Temam (1976, chapter III)) implies that

\[
\mu := \inf_{p \in H} \left[ J^*(K^* p) + J(u^\dagger) - \langle p, g \rangle + \frac{\| p \|^2}{2t} \right] = -\inf_{u \in X} \left[ \frac{t}{2} \| Ku - g \|^2 + J(u) - J(u^\dagger) \right],
\]

as the right-hand side of this equation is the dual of the left-hand side. Using the variational inequality (9) and the non-negativity of \( D \), we therefore find that

\[
\mu \leq \sup_{u \in X} \left[ \Phi(\| Ku - g \|^2) - D(u, u^\dagger) - \frac{t}{2} \| Ku - g \|^2 \right]
\leq \sup_{u \in X} \Phi(\| Ku - g \|^2) - \frac{t}{2} \| Ku - g \|^2.
\]

Replacing \( \| Ku - g \|^2 \) by \( s \geq 0 \) in the last term and using the definition of \( \Psi \), we obtain

\[
\mu \leq \sup_{s \geq 0} \left[ \Phi(s) - \frac{ts}{2} \right] = \frac{t}{2} \sup_{s \geq 0} \left[ \frac{2\Phi(s)}{t} - s \right] = \frac{t}{2} \sup_{s \geq 0} \left[ \frac{2s}{t} - \Phi^{-1}(s) \right] = \frac{t}{2} \psi \left( \frac{2}{t} \right),
\]

which proves the assertion.

We close this section with a statement concerning the dual variables \( \{ p^\delta_1, p^\delta_2, \ldots \} \) generated by the ALM. It is well known (in the case when \( \delta = 0 \)) that these stay bounded if and only if the source condition (7) holds. Assumption 3.2, however, allows control of their growth, as the following result shows.

Corollary 3.6. Let assumptions 2.1 and 3.2 hold. Then, there exists a constant \( C > 0 \), such that

\[
\| p^\delta_n \|^2 \leq Ct_n^2 \left( \psi \left( \frac{2}{t_n} \right) + \delta^2 \right).
\]

Proof. It follows from (20) that there exists a constant \( C > 0 \), such that

\[
\| p^\delta_n \|^2 \leq Ct_n \left( J^*(K^* p) + J(u^\dagger) - \langle p, g \rangle + \frac{\| p \|^2}{2t_n} + t_n \delta^2 \right)
\]

for all \( p \in H \). Applying lemma 3.5 yields the desired estimate.
4. Morozov’s discrepancy principle

In this section, we study Morozov’s discrepancy principle as an \textit{a posteriori} stopping rule for the ALM. To be more precise, if \([u^1, u^2, \ldots] \) is generated by the ALM, Morozov’s rule suggests stopping the iteration at the index

\[ n^*(\delta) = \min \{ n \in \mathbb{N} : \| Ku^*_n - g \delta \| \leq \rho \delta \}, \quad (22) \]

where \( \rho > 1 \). In this section, we prove convergence rates for the iterates \( u^*_n(\delta) \) given that assumption 3.2 holds. Morozov’s principle for the case when the source condition (7) holds is trivial. Denote for the sake of simplicity \( \bar{\rho} \). Now assume that

\[ \| g \| \leq \bar{\rho} \delta \]

implies in particular the inequality

\[ \| Ku_n - g \| \leq \rho \delta \]

for some \( \rho > 0 \) independent of \( n \), such that

\[ D(u^*_n(\delta), u^\dagger) \leq \frac{C(\rho + 1)^2 \delta^2}{\Psi^{-1}(\rho^2 - 1) \delta^2} + C(\rho + 1)^2 \delta^2 \sup_{k \in \mathbb{N}} t_k. \]

\textbf{Remark 4.3.} Assume that the variational inequality (9) is satisfied with \( \Phi(s) = C \delta^p \) for some \( C > 0 \) and \( p > 0 \). Then, setting \( u = u^\dagger + tz \) for some \( z \in X \) and \( t > 0 \), the non-negativity of \( D \) implies in particular the inequality

\[ J(u^\dagger) - J(u^\dagger + t z) \leq C t^p \| Kz \|^p. \]

Now assume that \( p > 1/2 \). Then, we obtain, after dividing by \( t \) and considering the limit \( t \to 0^+ \), that the directional derivative of \( J \) satisfies \( -J'(u^\dagger)(z) \leq 0 \). Because \( z \) was arbitrary, this implies that \( u^\dagger \) minimizes the regularization term \( J \). Thus, the variational inequality can hold in non-trivial situations, if and only if \( p \leq 1/2 \).

Now note that the same condition is required for the function \( \Phi(s)^2/s = C t^2 s^{2p-1} \) to be non-increasing. Therefore, in the case of a variational inequality of Hölder type, assumption 4.1 imposes no relevant further restrictions on the index function.

Before we give the proof of theorem 4.2, we state the following lemma, which is interesting in its own right.

\textbf{Lemma 4.4.} Let assumptions 2.1 and 3.2 hold and assume that \( n^*(\delta) \) is chosen according to Morozov’s discrepancy principle (22). Then,

\[ t^*_n(\delta) \leq \frac{2}{\Psi^{-1}(\rho^2 - 1) \delta^2} + t^*_n(\delta). \]

\textbf{Proof.} Without loss of generality, we may assume that \( n^*(\delta) > 1 \); otherwise, the assertion is trivial. Denote for the sake of simplicity \( \bar{n} := n^*(\delta) - 1 \). Then, it follows from (22) that

\[ \| Ku^*_n - g \| \leq \rho \delta \]

Plugging this relation into (17) yields

\[ \rho^2 t_\bar{n} \delta^2 = \frac{\| p - p^\bar{n} \|^2}{2t_\bar{n}} + \frac{\| p - p^\bar{n} \|^2}{2t_\bar{n}} \]

\[ \leq J(K^* p) - (p, g) + J(u^\dagger) + \frac{\| p \|^2}{2t_\bar{n}} + \{ p - p^\bar{n}, g - g^\dagger \}. \]
for every $p \in H$. Applying Young’s inequality
\[
\langle p - p_\delta^s, g - g\rangle \leq \frac{\|p - p_\delta^s\|^2}{2t_\delta} + t_\delta \delta^2,
\]
we obtain with lemma 3.5 the estimate
\[
\frac{(\rho^2 - 1)t_\delta \delta^2}{2} \leq \inf_{p \in H} \left[ J^*(K^* p) - \langle p, g \rangle + J(u^t) + \frac{\|p\|^2}{2t_\delta} \right] \leq \frac{t_\delta}{2} \Psi\left(\frac{2}{t_\delta}\right).
\]
This proves that $(\rho^2 - 1)\delta^2 \leq \Psi(2/t_\delta)$. Now the assertion follows by applying the monotonically increasing function $\Psi^{-1}$ to both sides of this inequality and adding the last step size $\tau_{n^*(\delta)}$. □

Next, we need another lemma, which relates the condition on $\Phi$ in assumption 4.1 to an equivalent condition on the function $\Psi = (\Phi^{-1})^\ast$.

**Lemma 4.5.** Let $\Phi$ be an index function and $\Psi$ the Fenchel conjugate of $\Phi^{-1}$. Then, the mapping $s \mapsto \Phi(s)^{2}/s$ is non-increasing, if and only if the mapping $t \mapsto t^2\Psi(2/t)$ is non-decreasing.

**Proof.** First note that, by means of the change of variables $t \mapsto 2/t$ and ignoring the constant factor, the mapping $t \mapsto t^2\Psi(2/t)$ is non-decreasing, if and only if the mapping $t \mapsto H(t) := \Psi(t)/t^2$ is non-decreasing. Because $\Psi$ is convex and continuous, this condition is satisfied, if and only if $H'(t) \leq 0$ for every $t \geq 0$ for which $\Psi'(t)$ exists. Now,
\[
H'(t) = \frac{\Psi'(t)}{t^3} - \frac{2\Psi(t)}{t^3} = \frac{1}{t^3}(t\Psi'(t) - 2\Psi(t)),
\]
and therefore, $H'(t) \leq 0$ if and only if $t\Psi'(t) - 2\Psi(t) \leq 0$. Now recall that $\Psi$ is the Fenchel conjugate of $\Phi^{-1}$, and therefore, $t\Psi'(t) = \Psi(t) + \Phi^{-1}(\Psi'(t))$. Thus, $H'(t) \leq 0$, if and only if $\Phi^{-1}(\Psi'(t)) - \Psi(t) \leq 0$.

Similarly, the mapping $s \mapsto \Phi(s)^{2}/s$ is non-increasing, if and only if the mapping $s \mapsto \tilde{H}(s) := s^2/\Phi^{-1}(s)$ is non-increasing, which in turn is equivalent to the condition
\[
\tilde{H}'(s) = \frac{2s}{\Phi^{-1}(s)} - \frac{s^2\Phi^{-1'}(s)}{\Phi^{-1}(s)^2} = \frac{s(2\Phi^{-1}(s) - s\Phi^{-1'}(s))}{\Phi^{-1}(s)^2} \leq 0.
\]
Because of the equality $s\Phi^{-1'}(s) = \Phi^{-1}(s) + \Psi(\Phi^{-1'}(s))$, this is the case, if and only if $\Phi^{-1}(s) - \Psi(\Phi^{-1'}(s)) \leq 0$. The assertion now follows from the fact that $s = \Psi'(t)$ if and only if $t = \Phi^{-1}(s)$, which, again, is a consequence of the fact that $\Phi^{-1}$ and $\Psi$ are conjugate. □

**Proof of theorem 4.2.** Throughout the proof, we use the abbreviation $n = n^*(\delta)$. First, observe that $K^* p_\delta^s \in \partial J(u_\delta^s)$, and thus, $J(u_\delta^s) - J(u^t) \leq \langle p_\delta^s, Ku_\delta^s - g \rangle$. From the discrepancy rule (22), it follows that
\[
\|Ku_\delta^s - g\| \leq \|Ku_\delta^s - g^\delta\| + \delta \leq (\rho + 1)\delta,
\]
and hence, the variational inequality (9) implies
\[
D(u_\delta^s, u^t) \leq \|p_\delta^s\|(\rho + 1)\delta + \Phi((\rho + 1)^2\delta^2).
\]
(23)

As in the proof of lemma 3.4, we observe that for all $s, r \geq 0$ one has $s\Phi(r) \leq \Psi(s) + r$. Setting $r = (\rho + 1)^2\delta^2$ and $s = \Psi^{-1}((\rho^2 - 1)\delta^2)$, one finds, after dividing both sides of the inequality by $s$, that
\[
\Phi((\rho + 1)^2\delta^2) \leq \frac{(\rho^2 - 1)\delta^2}{\Psi^{-1}((\rho^2 - 1)\delta^2)} + \frac{(\rho + 1)^2\delta^2}{\Psi^{-1}((\rho^2 - 1)\delta^2)} = \frac{2\rho(\rho + 1)\delta^2}{\Psi^{-1}((\rho^2 - 1)\delta^2)}.
\]
which yields an estimate for the second term in (23). For estimating the first term, we note that corollary 3.6 implies the estimate
\[
\|p_n\| \leq \tilde{C} t_n \left( \psi \left( \frac{2}{t_n} \right) + \delta^2 \right)^{1/2} \leq \tilde{C} t_n \psi \left( \frac{2}{t_n} \right)^{1/2} + \tilde{C} t_n \delta
\]
for some constant \( \tilde{C} > 0 \). By assumption, the mapping \( x \mapsto \Phi(x^2/x) \) is non-increasing, and therefore, using lemma 4.5, the mapping \( s \mapsto s^2 \psi(2/s) \) is non-decreasing. Thus, we obtain, after using the estimate for \( t_n \) of lemma 4.4 and the monotonicity of \( \psi \),
\[
\|p_n\| \leq \tilde{C} \left( \frac{2}{\psi^{-1}((\rho^2 - 1)\delta^2)} + \tau_n \right) \psi \left( \frac{2\psi^{-1}((\rho^2 - 1)\delta^2)}{2 + \tau_n \psi^{-1}((\rho^2 - 1)\delta^2)} \right)^{1/2} + \tilde{C} \tau_n \delta
\]
Consequently, we have
\[
D(u_0^\dagger, u^\dagger) \leq \frac{2(\tilde{C}(\rho + 1)^2\delta^2)}{\psi^{-1}((\rho^2 - 1)\delta^2)} + \tilde{C}(\rho + 1)^2\tau_n \delta^2 + \frac{2\rho(\rho + 1)\delta^2}{\psi^{-1}((\rho^2 - 1)\delta^2)} \leq 2(\tilde{C} + 1)(\rho + 1)^2\delta^2 \sup_{\xi} \tau_k
\]
which proves the assertion with \( C := 2(\tilde{C} + 1) \). \( \Box \)

5. Examples

In this section, we discuss particular instances of the variational inequality (9) and the implications of the general results in sections 3 and 4 for these special scenarios. The first two examples shed some light on the relation of variational inequalities and more standard notions of source conditions: the KKT condition (7) and Hölder-type conditions. The third example deals with sparsity promoting regularization, where standard notions of source conditions together with an additional restricted injectivity assumption allow the derivation of convergence rates with respect to the norm instead of the Bregman distance.

5.1. Standard source condition

It is quite easy to see that the standard source condition (7) implies the variational inequality (11). Indeed, assume that \( u^\dagger \) is a solution of (1) and that \( K^* p^\dagger \in \partial J(u^\dagger) \) for some \( p^\dagger \in H \). By defining \( \xi^\dagger = K^* p^\dagger \), one observes
\[
\langle \xi^\dagger, u^\dagger - u \rangle_{X^*,X} = \langle p^\dagger, g - Ku \rangle \leq \| p^\dagger \| \| Ku - g \|.
\]
Setting \( \beta = 1 \) and \( \Phi(t) = \| p^\dagger \| t^{1/2} \) gives (11).

The converse is in general not true, i.e. (11) with \( \Phi(t) = \gamma t^{1/2} (\gamma > 0) \) does not imply the existence of \( p^\dagger \in V \), such that \( K^* p^\dagger \in \partial J(u^\dagger) \). However, if (11) is replaced by the stronger condition
\[
\langle \xi^\dagger, u^\dagger - u \rangle_{X^*,X} \leq (1 - \beta) D_J(u, u^\dagger) + \gamma \| Ku - g \|, \quad (24)
\]
for all \( u \in X \), then the two notions are equivalent. Here, \( D_J(u, v) = J(u) - J(v) - J'(v)(u - v) \)
and \( J'(v)(w) \) is the directional derivative of \( J \) at \( v \) in the direction \( w \):
\[
J'(v)(w) = \lim_{h \to 0^+} \frac{1}{h} (J(v + hw) - J(v)).
\]

Note that, for convex \( J \), the directional derivative is well defined for every \( v \) and \( w \) (though it takes values in \([−\infty, \infty]\)) and is positively one-homogeneous, i.e. \( J'(v)(tw) = tJ'(v)(w) \) for all \( t > 0 \).

In order to see the aforementioned equivalence, let \( v \in X \) and set \( u = u^i - tv \) in (24) for
some \( t > 0 \). Then,
\[
(\xi^+, tv)^*_X \leq (1 - \beta)D_J(u^i - tv, u^i) + \gamma \|tv\|.
\]

Since the mapping \( w \mapsto J'(u^i)(w) \) is positively one-homogeneous, this implies that
\[
(\xi^+, v)^*_X \leq (1 - \beta)J(u^i - tv) - J(u^i) + \gamma \|tv\|,
\]
for all \( v \in X \) and \( t > 0 \). Letting \( t \to 0^+ \), this shows that \( (\xi^+, v)^*_X \leq \gamma \|v\| \) for all \( v \in X \),
and hence, \( K^p \rho = \xi^+ \) for some \( \rho \in H \) according to Scherzer et al (2009, lemma 8.21).

In the particular case where the mapping \( J \) is Gâteaux differentiable at \( u^i \), the
subdifferential \( dJ(u^i) \) contains a single element \( \xi^+ \), which coincides with the directional
derivative, i.e., \( (\xi^+, v) = J'(u^i)(v) \) for every \( v \in X \). Thus, in this case, the source condition
is equivalent with the variational inequality.

If \( \Phi(t) = \gamma t^{1/2} \), then the Fenchel conjugate \( \Psi \) of \( \Phi^{-1} \) reads \( \Psi(t) = \gamma/(2\sqrt{2})t^2 \). Hence,
it follows from theorem 3.3 that there exists a constant \( C > 0 \), such that
\[
D_J^{K^p} (u^i, u^i) \leq C\delta
\]
given the a priori stopping rule \( t_n \asymp \delta^{-1} \). This is the well-known convergence rate result
for the standard source condition (see Burger et al (2007), Frick and Scherzer (2010)).
We note that the results by Frick and Scherzer (2010) are slightly stronger, as they give \( \delta \)-rates
for the symmetric Bregman distance (see also Frick et al (2011)). If Morozov’s discrepancy
principle (22) is applied as an a posteriori stopping rule, we obtain from theorem 4.2 that
\[
D_J^{K^p} (u^i, u^i) \leq C\sqrt{\frac{(\rho + 1)^3}{\rho - 1}} \delta + C(\rho + 1)^2\delta^2 \sup_{k \in \mathbb{N}} t_k.
\]
This coincides with the results by Frick et al (2011, theorem 4.3), where Morozov’s discrepancy
rule for the standard source condition was studied.

5.2. Hölder-type conditions

In this section, we study the relationship between the variational inequality (11) and Hölder-type
source conditions for the iteration (4).

We first consider the case of the iterated Tikhonov method, i.e., \( L = \text{Id} \) and thus \( J(u) = \|u\|^2/2 \).
Then, a solution \( u^i \) of (1) is said to satisfy a Hölder condition with exponent
\( 0 \leq v < \frac{1}{2} \) if \((K^*K)^{v}\rho = u^i = dJ(u^i)\). If \( u^i \) satisfies a Hölder condition
with exponent \( v \), then (11) holds with \( D_J^{v} (u^i, u^i) = \|u - u^i\|^2 \) and \( \Phi(s) = s^{1+1/2v} \). To see this, observe, that the
interpolation inequality (cf Engl et al (1996, p 47)) implies
\[
\langle u^i, u^i - u \rangle \leq \|p\| \|(K^*K)^{\frac{1}{2}}(u^i - u)\| \leq \|p\| \|(K^*K)^{\frac{1}{2}}(u^i - u)\|^{2v} \|u^i - u\|^{1-2v} = 2^{1-v} \|p\| \|(Ku - g)\|^v \|D_J^{v} (u, u^i)\|^{\frac{2}{1-2v}}.
\]
Using Young’s inequality $ab \leq a^\alpha p + b^\beta q$ with $q = 2/(1 - 2\nu)$ and $p = 2/(1 + 2\nu)$ shows for all $\eta > 0$

$$
(\|Ku - g\|^2)D^\nu_f(u, u^\dagger) \leq \frac{1}{\eta} (\|Ku - g\|^2)\eta D^\nu_f(u, u^\dagger) + \frac{1 + 2\nu}{2\eta} (\|Ku - g\|^2)^{\frac{1}{\nu}} + \frac{\eta^{\frac{1}{2\nu}} (1 - 2\nu)D^\nu_f(u, u^\dagger)}{2}.
$$

Choosing $\eta$ such that $1 - \beta = \eta^{\frac{1}{2\nu}} \|p^!\| (\frac{1 - 2\nu}{2})^{\frac{1}{2\nu}} < 1$ results in (11) after setting $\Phi(s) = c \eta^{\frac{1}{2\nu}}$ with $c = \frac{1 + 2\nu}{2\eta^{\frac{1}{2\nu}}}$.

In the case of the \emph{iterated Tikhonov–Morozov method}, we consider (4) with $K = \text{Id}$ and $L : D(L) \subset X \rightarrow H$ being a densely defined, closed linear operator. Recall that in this case, $\hat{L} = (\text{Id} + LL^*)^{-1}$ and $\hat{L} = (\text{Id} + L^*L)^{-1}$ are self-adjoint and bounded linear operators (cf. Groetsch (2007, chapter 2.4)). A solution $u^\dagger$ of (1) is said to satisfy a Hölder condition with the exponent $0 \leq \nu \leq \frac{1}{2}$ if $Lu^\dagger = \hat{L}^\nu$ for some $\omega^\nu \in H$. We show that this condition implies \eqref{eqn:9} when $D(u, u^\dagger)$ equals $\|Lu - Lu^\dagger\|^2$ (for some $\gamma \in (0, 1)$) whenever $u \in D(L)$ and $+\infty$ else. To see this, recall that $J(u) = \infty$ if $u \notin D(L)$. Thus, \eqref{eqn:9} is equivalent to

$$
\langle Lu^\dagger, Lu^\dagger - Lu \rangle \leq (1 - \gamma) \|Lu - Lu^\dagger\|^2 + \Phi(\|u - u^\dagger\|^2)
$$

(25)

for all $u \in D(L)$. Setting $Lu^\dagger = \hat{L}^\nu \omega^\nu$ shows together with the interpolation inequality (Groetsch 2007, lemma 2.10) that for all $u \in D(L)$

$$
\langle Lu^\dagger, Lu^\dagger - Lu \rangle = \langle \omega^\nu, \hat{L}^\nu (Lu^\dagger - Lu) \rangle \\
\leq \|\omega^\nu\|\|\hat{L}^\nu (Lu^\dagger - Lu)\|^{2\nu} \|Lu^\dagger - Lu\|^{1 - 2\nu} \\
\leq \|\omega^\nu\|\|\hat{L}\|^{2\nu} \|u^\dagger - u\|^{2\nu} \|Lu^\dagger - Lu\|^{1 - 2\nu}.
$$

With the same arguments as in the case of the iterated Tikhonov method above, we conclude that (25) holds with $\Phi(s) = c \hat{c} s^{\frac{2\nu}{\nu - 1}}$ for some constant $\hat{c} > 0$.

Now, let $X$ again be a general Banach space and $J : X \rightarrow \mathbb{R}$ be convex such that assumptions 2.1 are satisfied. As revealed by the calculations above, the variational inequality \eqref{eqn:9} with $\Phi(s) \asymp s^{\frac{2\nu}{\nu - 1}}$ can be seen as a generalized Hölder condition. Note that in this case the Legendre conjugate $\Psi$ of $\Phi^{-1}$ behaves as $\Psi(t) \asymp t^{1 + 2\nu}$, and thus theorem 3.3 amounts to saying that there exists a constant $C > 0$, such that

$$
13
$$

$$
D(u^\dagger) \leq C \delta^{\frac{4\nu}{\nu - 1}}
$$

if $t_n \asymp \delta^{\frac{4\nu}{\nu - 1}}$. Morozov’s discrepancy principle (22) then shows that

$$
D(u^\dagger) \leq C \left(\frac{\rho + 1}{\rho - 1}\right)^{\frac{1}{\nu}} \delta^{\frac{4\nu}{\nu - 1}} + C (\rho + 1)^2 \delta^2 \sup_{k \in I} \tau_k.
$$

These results coincide with the lower order rates for the iterated Tikhonov method (Hanke and Groetsch 1998) and iterated Tikhonov–Morozov method (Groetsch 2007).

5.3. Sparsity promoting regularization

We now discuss the application of the results derived in this paper to sparsity promoting regularization. To that end, we assume that $X$ is a Hilbert space with orthonormal basis $\{\phi_i : i \in \mathbb{N}\}$, and we consider the regularization term $J(u) := \sum_i |\langle \phi_i, u \rangle|^q$ for some $1 \leq q < 2$ (see Daubechies et al (2004)). Grasmair et al (2008) has shown that, for Tikhonov regularization, this setting allows the derivation of convergence rates of order $O(\delta^{q})$ with respect to the norm, if $u^\dagger$ satisfies the standard source condition $K^* p^\dagger \in \partial J(u^\dagger)$ for some
and additionally, a restricted injectivity condition holds. In the following, we will generalize these results to the ALM and source conditions of Hölder type.

Assume that there exists $0 < \nu \leq 1/2$, such that $(K^*K)^q p^j = \xi_i \in \partial J(u^i)$ and that $\text{supp}(u^i) := \{ i \in \mathbb{N} : \langle \phi_i, u \rangle \neq 0 \}$ is finite. If $q = 1$, assume in addition that the restriction of $K$ to span$\{\phi_i : i \in \text{supp}(x^i)\}$ is injective, and in case $q = 1$, assume that the restriction of $K$ to span$\{\phi_i : |\langle \phi_i, x^i \rangle| < 1\}$ is injective. We will show in the following that, under these assumptions, there exists a constant $C > 0$, such that (9) holds with $D(u, u^i) = C\|u^j - u\|^q$ and $\Phi(s) \asymp s^{\frac{q}{1-2\nu}}$ in case $q > 1$, and with $D(u, u^i) = C\|u^i - u\|$ and $\Phi(s) \asymp s^2$ for $q = 1$.

It has been shown by Grasmair et al. (2008, proofs of theorems 13 and 15) that the given assumptions imply the existence of constants $C_1, C_2 > 0$, such that

$$C_1 \|u^j - u\| \leq C_2 \| Ku - g \|^q + J(u) - J(u^j) - \langle \xi_i, u - u^j \rangle$$

for all $u \in X$. Applying the interpolation inequality to $\langle \xi_i, u - u^j \rangle$, we obtain, similarly as in section 5.2, the estimate

$$C_1 \|u^j - u\|^q \leq C_2 \| Ku - g \|^q + J(u) - J(u^j) + \| p^j \| \| Ku - g \|^{2\nu} \| u^j - u \|^{1-2\nu}.$$ 

Now, Young’s inequality with $p = q/(1-2\nu)$ and $p_\nu = q/(q-1+2\nu)$ shows that

$$C_1 \|u^j - u\|^q \leq C_2 \| Ku - g \|^q + J(u) - J(u^j)
+ \| p^j \| \frac{q-1+2\nu}{q} \eta^{\frac{q-1}{q}} \| Ku - g \|^{\frac{2\nu}{q-1}}.$$

Choosing $\eta > 0$, such that $C = C_1 - \| p^j \| \frac{q-1+2\nu}{q} \eta^{\frac{q-1}{q}} > 0$, and setting

$$\Phi(s) = C_2 s^{\frac{q}{1-2\nu}} + \| p^j \| \frac{q-1+2\nu}{q} \eta^{\frac{q-1}{q}} s^{\frac{2\nu}{q-1}}.$$

we obtain the variational inequality (9). Because $\frac{2\nu}{q-1-2\nu} \leq q$, the asymptotic behaviour of $\Phi$ for $s \to 0$ is governed by its second term, which shows that $\Phi(s) \asymp s^{\frac{q}{1-2\nu}}$. Moreover, in the special case $q = 1$, the term $s^{\frac{q}{1-2\nu}}$ reduces to $s^2$ independent of the type of the source condition. For the function $\Psi$, we obtain the asymptotic behaviour $\Psi(s) \asymp s^{\frac{q}{q-1-2\nu}}$. Thus, theorem 3.3 shows that for $t_n \asymp \delta^{-\frac{1}{4\nu+2\nu^2}}$, we have the estimate

$$\| u_n^j - u^j \| \leq C \delta^{\frac{2\nu}{q-1-2\nu}}$$

for $\delta > 0$ sufficiently small, and a similar estimate for Morozov’s discrepancy principle.

**Remark 5.1.** Grasmair et al. (2011) has shown for Tikhonov regularization with $\sum_i |\langle \phi_i, u \rangle|$, which is the special case of the ALM with a single-iteration step, that a linear convergence rate with respect to the norm is equivalent to the usual source condition. Thus, the above-mentioned results imply that, in the case $q = 1$, the Hölder-type source condition $(K^*K)^q p^j \in \partial J(u^i)$ in fact already implies the standard source condition $K^*p^j \in \partial J(u^i)$ for some different source element $p^j$.

**Acknowledgments**

The second author would like to thank Axel Munk and the staff of the Institute for Mathematical Stochastics, University of Göttingen for their hospitality during his stay in Göttingen. This work was partially funded by the DFG-SNF Research Group FOR916 Statistical Regularization and Qualitative Constraints (Z-Project).
References

Aubin J P 1979 Mathematical Methods of Game and Economic Theory (Studies in Mathematics and Its Applications vol 7) (Amsterdam: North-Holland)

Bachmayr M and Burger M 2009 Iterative total variation schemes for nonlinear inverse problems Inverse Problems 25 105004

Boq R I and Hofmann B 2010 An extension of the variational inequality approach for obtaining convergence rates in regularization of nonlinear ill-posed problems J. Integral Eqns Appl. 22 369–92

Brill M and Schock E 1987 Model Optimization in Exploration Geophysics (Berlin, 1986) (Theory and Practice Applied Geophysics vol 1) (Braunschweig: Vieweg) pp 13–37

Burger M and Osher S 2004 Convergence rates of convex variational regularization Inverse Problems 20 1411–21

Burger M, Resmerita E and He L 2007 Error estimation for Bregman iterations and inverse scale space methods in image restoration Computing 81 109–35

Daubechies I, Defrise M and De Mol C 2004 An iterative thresholding algorithm for linear inverse problems with a sparsity constraint Commun. Pure Appl. Math. 57 1413–57

Ekeland I and Temam R 1976 Convex Analysis and Variational Problems (Studies in Mathematics and Its Applications vol 1) (Amsterdam: North-Holland)

Engl H, Hanke M and Neubauer A 1996 Regularization of Inverse Problems (Mathematics and Its Applications) (Dordrecht: Kluwer)

Fortin M and Glowinski R 1983 Applications to the numerical solution of boundary value problems Augmented Lagrangian Methods (Studies in Mathematics and Its Applications vol 15) (Amsterdam: North-Holland) (translated from French by B Hunt and D C Spicer)

Frick K, Lorenzi D A and Resmerita E 2011 Morozov’s principle for the augmented Lagrangian method applied to linear inverse problems Multiscale Modeling Simul. 9 1528–48

Frick K and Scherzer O 2010 Regularization of ill-posed linear equations by the non-stationary augmented Lagrangian method J. Integral Eqns Appl. 22 217–57

Grasmair M 2010 Generalized Bregman distances and convergence rates for non-convex regularization methods Inverse Problems 26 115014

Grasmair M, Haltmeier M and Scherzer O 2008 Sparse regularization with ℓp penalty term Inverse Problems 24 055020

Grasmair M, Haltmeier M and Scherzer O 2011 Necessary and sufficient conditions for linear convergence of ℓ1-regularization Commun. Pure Appl. Math. 64 161–82

Groetsch C W 2007 Stable Approximate Evaluation of Unbounded Operators (Lecture Notes in Mathematics vol 1894) (Berlin: Springer)

Groetsch C W and Scherzer O 2000 Non-stationary iterated Tikhonov–Morozov method and third-order differential equations for the evaluation of unbounded operators Math. Methods Appl. Sci. 23 1287–300

Güler O 1991 On the convergence of the proximal point algorithm for convex minimization SIAM J. Control Optim. 29 403–19

Hante M and Groetsch C W 1998 Nonstationary iterated Tikhonov regularization J. Optim. Theory Appl. 98 37–53

Hestenes M R 1969 Multiplier and gradient methods J. Optim. Theory Appl. 4 303–20

Hofmann B, Kaltenbacher B, Pöschl C and Scherzer O 2007 A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators Inverse Problems 23 987–1010

Ito K and Kunisch K 2008 Lagrange Multiplier Approach to Variational Problems and Applications (Advances in Design and Control vol 15) (Philadelphia, PA: SIAM)

Jung M, Resmerita E and Vese L 2012 Dual norm based iterative methods for image restoration J. Math. Imaging Vis. 44 128–49

Krasnosel’skii M A 1960 Solution of equations involving adjoint operators by successive approximations Usp. Mat. Nauk. 15 161–5

Kreinov A V 1973 Solution of ill-posed problems by successive approximation methods Dokl. Akad. Nauk SSSR 210 20–2

Lardy L J 1975 A series representation for the generalized inverse of a closed linear operator Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Nat. 58 152–7

Martinet B 1970 Régularisation d’inéquations variationnelles par approximations successives Rev. Fr. Inform. Rech. Optim. 4 154–8

Morozov V A 1967 Choice of parameter for the solution of functional equations by the regularization method Dokl. Akad. Nauk SSSR 175 1225–8

Osher S, Burger M, Goldfarb D, Xu J and Yin W 2005 An iterative regularization method for total variation-based image restoration Multiscale Modeling Simul. 4 460–89
Powell M J D 1969 A method for nonlinear constraints in minimization problems *Optimization* (Symp. Univ. Keele, 1968) (London: Academic) pp 283–98

Rockafellar R T 1974 Augmented Lagrange multiplier functions and duality in nonconvex programming *SIAM J. Control* 12 268–85

Scherzer O, Grasmair M, Grossauer H, Haltmeier M and Lenzen F 2009 *Variational Methods in Imaging* (Applied Mathematical Sciences vol 167) (New York: Springer)