K. Frick[1] and M. Grasmair[2]
[1]Institute for Mathematical Stochastics
University of Göttingen
Goldschmidtstraße 7, 37077 Göttingen, Geermany
[2]Computational Science Center
University of Vienna
Nordbergstraße 15, 1090 Vienna, Austria

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

Regularization of Linear Ill-posed Problems by the Augmented Lagrangian Method and Variational Inequalities

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.

AMS classification scheme numbers: 65J20, 47A52;

Submitted to: *Inverse Problems*

1. Introduction

We aim for the solution of the problem

$$\inf_{u \in X} J(u) \quad \text{s.t.} \quad Ku = g, \tag{1}$$

where $K\colon X \to H$ is a linear and bounded mapping between a Banach space X and a Hilbert space H and where $J\colon X \to \overline{\mathbb{R}}$ 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^δ such that

$$\left\| g - g^\delta \right\| \le \delta \tag{2}$$

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 $p_0^\delta \in H$ and for $k = 1, 2, \ldots$, computes

$$u_k^\delta \in \operatorname*{argmin}_{u\in X}\left[\frac{\tau_k}{2}\left\|Ku-g^\delta\right\|^2 + J(u) - \left\langle p_{k-1}^\delta, Ku-g^\delta\right\rangle\right] \tag{3a}$$

$$p_k^\delta = p_{k-1}^\delta + \tau_k(g^\delta - Ku_k^\delta). \tag{3b}$$

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 in (Hestenes 1969, Powell 1969) (under the name *method of multipliers*) as a solution method for problems of type (1) in Euclidean space with *exact* right hand side g. Since then, the ALM was developed further in various directions; see e.g. (Fortin & Glowinski 1983, Ito & 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 shortly discussed in (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 $L\colon D(L)\subset X\to\tilde{H}$, 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 into

$$u_k^\delta = \operatorname*{argmin}_{u\in X}\left[\tau_k\left\|Ku-g^\delta\right\|^2 + \left\|L(u-u_{k-1}^\delta)\right\|_{\tilde{H}}^2\right]. \tag{4}$$

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

A generalization of the iteration in (4) for total-variation based image reconstruction has been established in (Osher et al. 2005) under the name *Bregman iteration* and convergence properties were studied in (Burger et al. 2007). In (Frick & Scherzer 2010) it was pointed out that the Bregman iteration and the iterated Tikhonov(–Morozov) method are special instances of the ALM as it is stated in (3a), and an improved convergence analysis was developed. In (Frick et al. 2011), Morozov's discrepancy principle (Morozov 1967) was studied for the ALM. The application of the ALM for the regularization of nonlinear operators has been considered in (Bachmayr & Burger 2009, 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 & Osher 2004)

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

Here $K^*\colon H\to X^*$ 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 δ 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 paper is organized as follows: In Section 2 we state basic assumptions and review tools from 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 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 as well as review basic notions and facts from convex analysis. We start by delimiting minimal functional analytic requirements.

Assumption 2.1. *(i) X is a separable Banach space with topological dual X^*. We denote the duality pairing of X and X^* by $\langle \xi, x\rangle_{X^*,X} = \xi(x)$.*

(ii) The operator $K: X \to H$ is linear and continuous.

(iii) The functional $J: X \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ is convex, lower semicontinuous and proper with nonempty domain $D(J) = \{u \in X \ : \ J(u) < \infty\}$.

(iv) For each $g \in H$ and $c > 0$ the set

$$\Lambda(g, c) = \Big\{u \in X \ : \ \|Ku - g\|^2 + J(u) \le c\Big\}$$

is sequentially weakly pre-compact in X.

For our analysis we will make extensive use of tools from convex analysis (here, we refer to (Aubin 1979, Ekeland & Temam 1976) as standard references). 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

$$J(u) \ge J(u_0) + \langle \xi, u - u_0\rangle_{X^*,X}, \qquad \text{for all } u \in X.$$

In this case, we call ξ 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 $g \in K(D(J))$ and that the iteration ($3a$) is well defined. The proof is analogous to (Frick & Scherzer 2010, Lem. 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 $J^*: X^* \to \overline{\mathbb{R}}$ 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))

$$\inf_{p\in H}\Big[J^*(K^*p) - \langle p, g\rangle\Big]. \tag{6}$$

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

$$K^* p^\dagger \in \partial J(u^\dagger) \quad \text{and} \quad K u^\dagger = g. \tag{7}$$

From an inverse problems perspective, these conditions are understood as *source conditions* (Burger & Osher 2004) that delimit a class of particular regular solutions $u^\dagger$ of (1) that can be reconstructed from noisy data at a certain rate depending on the noise level δ. 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 $u^\dagger \in X$ be a solution of (1). Then

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

Proof. Define a function $\Gamma : X \times H \to \overline{\mathbb{R}}$ by setting $\Gamma(u,p) = J(u)$ if $Ku = g + p$ and $G(u,p) = +\infty$ otherwise. According to (Ekeland & Temam 1976, Chap III. Prop. 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 $p(0) = J(u^\dagger) < \infty$ it remains to prove lower semicontinuity. Therefore, let $\{p_k\}_{k \in \mathbb{N}}$ 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, that is, $h(p_k) = \Gamma(u_k, p_k)$.

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 $\lim_{k' \to \infty} J(u_{k'}) = \liminf_{k \to \infty} J(u_k)$. Moreover, we observe that $\|Ku_k - g\|^2 = \|p_k\|^2$ is bounded, since $p_k \to 0$. Thus it follows from Assumption 2.1 that there exists a further subsequence $\{u_{k''}\}$ such that $u_{k''} \rightharpoonup \hat{u}$ for some $\hat{u} \in X$. This implies that $Ku_{k''} \rightharpoonup K\hat{u} = g$, and the lower semicontinuity and convexity of J finally proves that

$$\liminf_{k \to \infty} h(p_k) = \lim_{k' \to \infty} J(u_{k'}) = \liminf_{k'' \to \infty} J(u_{k''}) \geq J(\hat{u}) \geq J(u^\dagger) = h(0).$$

□

A relation similar to the duality relation between the optimization problems (1) and (6) can be established for the ALM: As first observed in (Rockafellar 1974), the dual sequence $\{p_0^\delta, p_1^\delta, \ldots\}$ generated by the ALM can be characterized by the *proximal point method (PPM)*. To be more precise, for all $k \geq 1$,

$$p_k^\delta = \operatorname*{argmin}_{p \in H} \left[\frac{1}{2} \left\| p - p_{k-1}^\delta \right\|^2 + \tau_k \left(J^*(K^* p) - \langle p, g^\delta \rangle \right) \right]. \tag{8}$$

The PPM was introduced by Martinet in (Martinet 1970) for minimizing a convex functional, which in the present situation is the dual functional (6). The sequence $\{p_k^\delta\}$ 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 $J^*(K^* p_k^\delta) - \langle p_k^\delta, g^\delta \rangle$ 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 side of the equation $Ku = g$ is given exactly, that is, $\delta = 0$. 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, for instance, the results in (Fortin & Glowinski 1983, Ito & 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 $\|g^\delta - g\| \le \delta$. 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 δ. 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*

$$\lim_{\delta \to 0} \delta^2 t_{n(\delta)} = 0 \qquad \text{and} \qquad \lim_{\delta \to 0} t_{n(\delta)} = +\infty\,,$$

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

Proof. See (Frick & Scherzer 2010, Theorem 5.3). □

The previous result does not include any estimate of the speed of the convergence of the approximate solutions $u^\delta_{n(\delta)}$ 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, Frick & 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 $\Phi\colon [0,\infty) \to [0,\infty)$, *i.e., a non-negative continuous function that is strictly increasing and concave with $\Phi(0) = 0$. Moreover, $D\colon X \times X \to [0,\infty]$ satisfies $D(u,u) = 0$ whenever $u \in X$, and $u^\dagger$ is a solution of (1) is such that*

$$D(u,u^\dagger) \le J(u) - J(u^\dagger) + \Phi(\|Ku - g\|^2) \quad \text{for all } u \in X. \tag{9}$$

We denote by Ψ the Legendre-Fenchel conjugate of Φ^{-1}.

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

$$D_J^\xi(u,v) = J(u) - J(v) - \langle \xi, u - v\rangle_{X^*,X} \tag{10}$$

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

$$\left\langle \xi^\dagger, u^\dagger - u\right\rangle_{X^*,X} \le (1-\beta) D_J^{\xi^\dagger}(u,u^\dagger) + \Phi(\|Ku - g\|^2) \tag{11}$$

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

The following theorem asserts that the 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 Φ).

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

$$D(u_n^\delta, u^\dagger) \le C t_n \left(\Psi\left(\frac{16}{t_n}\right) + \delta^2 \right) \tag{12}$$

and

$$\left\| K u_n^\delta - g^\delta \right\|^2 \le C \left(\Psi\left(\frac{16}{t_n}\right) + \delta^2 \right). \tag{13}$$

In particular, if $t_n \asymp \frac{1}{\Psi^{-1}(\delta^2)}$, then

$$D(u_n^\delta, u^\dagger) = \mathcal{O}\left(\frac{\delta^2}{\Psi^{-1}(\delta^2)}\right) \quad \text{and} \quad \left\| K u_n^\delta - g \right\|^2 = \mathcal{O}(\delta^2). \tag{14}$$

Proof. The theorem is a consequence of the two Lemmas below. Combining the estimates derived in Lemma 3.4 and Lemma 3.5, it follows that

$$D(u_n^\delta, u^\dagger) \le \tilde{C} t_n \left(\Psi\left(\frac{16}{t_n}\right) + \delta^2 + \frac{1}{2}\Psi\left(\frac{2}{t_n}\right) \right)$$

for some $\tilde{C} > 0$ and a similar estimte holds for $\left\| K u_n^\delta - g^\delta \right\|^2$. Now note that the fact that Φ is an index function implies that $\Psi = (\Phi^{-1})^*$ is convex, non-negative, and $\Psi(0) = 0$. As a consequence, Ψ is non-decreasing and sub-additive on $[0, +\infty)$. Thus the inequalities (12) and (13) follow with $C = \tilde{C} + 1/2$. Finally, the estimate (14) follows from the sub-additivity of Ψ. $\square$

Lemma 3.4. Let Assumptions 2.1 and 3.2 hold and define for $p \in H$, $t > 0$ and $\delta \ge 0$

$$\psi(p, t, \delta) = \left(t\Psi(16/t) + t\delta^2 + J^*(K^* p) + J(u^\dagger) - \langle p, g \rangle + \frac{\|p\|^2}{2t} \right).$$

Then, there exists a constant $C > 0$ such that

$$D(u_n^\delta, u^\dagger) \le C\psi(p, t_n, \delta) \quad \text{and} \quad \left\| K u_n^\delta - g^\delta \right\|^2 \le C \frac{\psi(p, t_n, \delta)}{t_n} \tag{15}$$

for all $p \in H$.

Proof. Without loss of generality we assume that $p_0^\delta = 0$ and we shall use the abbreviation $G(p, g) = J^*(K^* p) - \langle p, g \rangle$. In (Güler 1991, Lem. 2.1) it was proved that for all $p \in V$

$$\frac{t_n \left\| p_n^\delta - p_{n-1}^\delta \right\|^2}{2\tau_n^2} \le G(p, g^\delta) - G(p_n^\delta, g^\delta) - \frac{\left\| p - p_n^\delta \right\|^2}{2t_n} + \frac{\|p\|^2}{2t_n}. \tag{16}$$

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 $p_n^\delta - p_{n-1}^\delta = \tau_n(g^\delta - K u_n^\delta)$, this implies that

$$\begin{aligned}
\frac{t_n}{2}\left\|Ku_n^\delta - g^\delta\right\|^2 &\le G(p,g) - G(p_n^\delta, g) - \frac{\left\|p - p_n^\delta\right\|^2}{2t_n} + \frac{\|p\|^2}{2t_n} + \left\langle p - p_n^\delta, g - g^\delta\right\rangle \\
&\le G(p,g) + J(u^\dagger) - \frac{\left\|p - p_n^\delta\right\|^2}{2t_n} + \frac{\|p\|^2}{2t_n} + \left\langle p - p_n^\delta, g - g^\delta\right\rangle, \qquad (17)
\end{aligned}$$

where the second inequality follows from Lemma 2.2. Setting $p = p_n^\delta$, this proves that

$$\frac{t_n}{2}\left\|Ku_n^\delta - g^\delta\right\|^2 \le J^*(K^* p_n^\delta) - \left\langle p_n^\delta, g\right\rangle + J(u^\dagger) + \frac{\left\|p_n^\delta\right\|^2}{2t_n}.$$

Since $K^* p_n^\delta \in \partial J(u_n^\delta)$, we observe that $J^*(K^* p_n^\delta) + J(u_n^\delta) = \left\langle K^* p_n^\delta, u_n^\delta\right\rangle$ and conclude that

$$\begin{aligned}
\frac{t_n}{2}\left\|Ku_n^\delta - g^\delta\right\|^2 &\le J(u^\dagger) - J(u_n^\delta) + \left\langle p_n^\delta, Ku_n^\delta - g\right\rangle + \frac{\left\|p_n^\delta\right\|^2}{2t_n} \\
&= J(u^\dagger) - J(u_n^\delta) + \left\langle p_n^\delta, Ku_n^\delta - g^\delta\right\rangle + \left\langle p_n^\delta, g^\delta - g\right\rangle + \frac{\left\|p_n^\delta\right\|^2}{2t_n}.
\end{aligned}$$

Applying Young's inequality $\langle a, b\rangle \le \|a\|^2/2 + \|b\|^2/2$ 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

$$\begin{aligned}
\frac{t_n}{4}\left\|Ku_n^\delta - g^\delta\right\|^2 &\le J(u^\dagger) - J(u_n^\delta) + \left\langle p_n^\delta, g^\delta - g\right\rangle + \frac{3\left\|p_n^\delta\right\|^2}{2t_n} \\
&\le J(u^\dagger) - J(u_n^\delta) + \frac{\delta^2 t_n}{2} + \frac{2\left\|p_n^\delta\right\|^2}{t_n}.
\end{aligned}$$

Summarizing, we find that

$$\left\|Ku_n^\delta - g^\delta\right\|^2 \le \frac{4}{t_n}\left(J(u^\dagger) - J(u_n^\delta)\right) + 2\delta^2 + \frac{8\left\|p_n^\delta\right\|^2}{t_n^2}.$$

Now, we observe from (9) that $J(u^\dagger) - J(u_n^\delta) \le -D(u_n^\delta, u^\dagger) + \Phi(\left\|Ku_n^\delta - g\right\|^2)$. Plugging this inequality into the above estimate yields

$$\left\|Ku_n^\delta - g^\delta\right\|^2 + \frac{4}{t_n}D(u_n^\delta, u^\dagger) \le \frac{4}{t_n}\Phi(\left\|Ku_n^\delta - g\right\|^2) + 2\delta^2 + \frac{8\left\|p_n^\delta\right\|^2}{t_n^2}. \qquad (18)$$

Since Ψ is the Legendre-Fenchel conjugate of $t \mapsto \Phi^{-1}(t)$, i.e., $\Psi(s) = \sup_{t\ge 0} st - \Phi^{-1}(t)$, it follows that $st \le \Psi(s) + \Phi^{-1}(t)$ for all $s, t \ge 0$, and in particular, for $t = \Phi(r)$, that $s\Phi(r) \le \Psi(s) + r$ for all $s, r \ge 0$. Setting $s = 16/t_n$ and $r = \left\|Ku_n^\delta - g^\delta\right\|^2$ gives

$$\begin{aligned}
\frac{4}{t_n}\Phi(\left\|Ku_n^\delta - g\right\|^2) &= \frac{1}{4}\frac{16}{t_n}\Phi(\left\|Ku_n^\delta - g\right\|^2) \\
&\le \frac{1}{4}\Psi\left(\frac{16}{t_n}\right) + \frac{1}{4}\left\|Ku_n^\delta - g\right\|^2 \\
&\le \frac{1}{4}\Psi\left(\frac{16}{t_n}\right) + \frac{1}{2}\left\|Ku_n^\delta - g^\delta\right\|^2 + \frac{\delta^2}{2}.
\end{aligned}$$

Combining this with (18) yields

$$\frac{1}{2}\left\|Ku_n^\delta - g^\delta\right\|^2 + \frac{4}{t_n}D(u_n^\delta, u^\dagger) \le \frac{1}{4}\Psi\left(\frac{16}{t_n}\right) + \frac{5\delta^2}{2} + \frac{8\left\|p_n^\delta\right\|^2}{t_n^2}. \tag{19}$$

Finally, we observe again from (16) that for all $p \in H$

$$\begin{aligned}
\frac{\left\|p - p_n^\delta\right\|^2}{2t_n^2} &\le \frac{G(p, g^\delta) - G(p_n^\delta, g^\delta)}{t_n} + \frac{\|p\|^2}{2t_n^2} \\
&\le \frac{G(p, g) - G(p_n^\delta, g)}{t_n} + \frac{1}{t_n}\left\langle p - p_n^\delta, g - g^\delta\right\rangle + \frac{\|p\|^2}{2t_n^2} \\
&\le \frac{G(p, g) - \inf_{q\in V} G(q, g)}{t_n} + \frac{\left\|p - p_n^\delta\right\|^2}{4t_n^2} + \delta^2 + \frac{\|p\|^2}{2t_n^2}.
\end{aligned}$$

This shows that

$$\begin{aligned}
\frac{\left\|p_n^\delta\right\|^2}{8t_n^2} &\le \frac{\left\|p - p_n^\delta\right\|^2}{4t_n^2} + \frac{\|p\|^2}{4t_n^2} && (20)\\
&\le \frac{G(p, g) - \inf_{q\in V} G(q, g)}{t_n} + \delta^2 + \frac{3\|p\|^2}{4t_n^2}. && (21)
\end{aligned}$$

Combining (20) with (19) and applying Lemma 2.2 finally gives

$$\begin{aligned}
\frac{1}{2}\left\|Ku_n^\delta - g^\delta\right\|^2 &+ \frac{4}{t_n}D(u_n^\delta, u^\dagger) \le \frac{1}{4}\Psi\left(\frac{16}{t_n}\right) + \frac{5\delta^2}{2} + \frac{8\left\|p_n^\delta\right\|^2}{t_n^2} \\
&\le \frac{1}{4}\Psi\left(\frac{16}{t_n}\right) + 64\frac{G(p, g) - \inf_{q\in V} G(q, g)}{t_n} + \frac{133\delta^2}{2} + \frac{48\|p\|^2}{t_n^2}.
\end{aligned}$$

□

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

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

Proof. Classical duality theory (see (Ekeland & Temam 1976, Chap 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

$$\begin{aligned}
\mu &\le \sup_{u\in X}\left[\Phi\big(\|Ku - g\|^2\big) - D(u, u^\dagger) - \frac{t}{2}\|Ku - g\|^2\right] \\
&\le \sup_{u\in X}\left[\Phi\big(\|Ku - g\|^2\big) - \frac{t}{2}\|Ku - g\|^2\right].
\end{aligned}$$

Replacing $\|Ku - g\|^2$ by $s \ge 0$ in the last term and using the definition of Ψ, we obtain

$$\mu \le \sup_{s\ge 0}\left[\Phi(s) - \frac{ts}{2}\right] = \frac{t}{2}\sup_{s\ge 0}\left[\frac{2\Phi(s)}{t} - s\right] = \frac{t}{2}\sup_{s\ge 0}\left[\frac{2s}{t} - \Phi^{-1}(s)\right] = \frac{t}{2}\Psi\left(\frac{2}{t}\right),$$

which proves the assertion. □

We close this section with a statement concerning the dual variables $\{p_1^\delta, p_2^\delta, \dots\}$ 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

$$\left\|p_n^\delta\right\|^2 \le C t_n^2 \left(\Psi\left(\frac{2}{t_n}\right) + \delta^2\right)$$

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

$$\left\|p_n^\delta\right\|^2 \le C t_n \left(J^*(K^* p) + J(u^\dagger) - \langle p, g\rangle + \frac{\|p\|^2}{2t_n} + t_n \delta^2\right)$$

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 a posteriori stopping rule for the ALM. To be more precise, if $\{u_1^\delta, u_2^\delta, \dots\}$ is generated by the ALM, Morozov's rule suggests stopping the iteration at the index

$$n^*(\delta) = \min\left\{n \in \mathbb{N} \ : \ \left\|K u_n^\delta - g^\delta\right\| \le \rho\delta\right\}, \tag{22}$$

where $\rho > 1$. In this section we prove convergence rates for the iterates $u_{n^*(\delta)}^\delta$ given that Assumption 3.2 holds. Morozov's principle for the case when the source condition (7) holds was studied in (Frick et al. 2011). Theorem 4.2 below extends this result to regularity classes that are delimited by the variational inequality in Assumption 3.2. Additionally to these, we will assume

Assumption 4.1. *Let Assumption 3.2 hold.*

(i) The mapping $s \mapsto \Phi(s)^2/s$ is non-increasing.

(ii) The sequence of stepsizes $\{\tau_1, \tau_2, \dots\}$ in the ALM is bounded.

Theorem 4.2. *Let Assumptions 2.1 and 4.1 hold and assume that $n^*(\delta)$ is chosen according to Morozov's discrepancy principle (22) for some $\rho > 1$. Then there exists a constant $C > 0$ independent of ρ such that*

$$D(u_{n^*(\delta)}^\delta, u^\dagger) \le \frac{C(\rho+1)^2\delta^2}{\Psi^{-1}\big((\rho^2-1)\delta^2\big)} + C(\rho+1)^2\delta^2 \sup_{k\in\mathbb{N}} \tau_k .$$

Remark 4.3. Assume that the variational inequality (9) is satisfied with $\Phi(s) = Cs^p$ 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

$$J(u^\dagger) - J(u^\dagger + tz) \le C t^{2p} \|Kz\|^{2p} .$$

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 $-J'(u^\dagger)(z) \le 0$. 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 $p \le 1/2$.

Now note that the same condition is required for the function $\Phi(s)^2/s = C^2 s^{2p-1}$ 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.

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,

$$t_{n^*(\delta)} \le \frac{2}{\Psi^{-1}((\rho^2-1)\delta^2)} + \tau_{n^*(\delta)}.$$

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 $\left\|Ku_{\bar{n}}^\delta - g^\delta\right\|^2 > \rho^2\delta^2$. Plugging in this relation into (17) yields

$$\begin{aligned} \frac{\rho^2 t_{\bar{n}}\delta^2}{2} + \frac{\left\|p - p_{\bar{n}}^\delta\right\|^2}{2t_{\bar{n}}} &< \frac{t_{\bar{n}}}{2}\left\|Ku_{\bar{n}}^\delta - g^\delta\right\|^2 + \frac{\left\|p - p_{\bar{n}}^\delta\right\|^2}{2t_{\bar{n}}} \\ &\le J^*(K^*p) - \langle p, g\rangle + J(u^\dagger) + \frac{\|p\|^2}{2t_{\bar{n}}} + \left\langle p - p_{\bar{n}}^\delta, g - g^\delta\right\rangle \end{aligned}$$

for every $p \in H$. Applying Young's inequality

$$\left\langle p - p_{\bar{n}}^\delta, g - g^\delta\right\rangle \le \frac{\left\|p - p_{\bar{n}}^\delta\right\|^2}{2t_{\bar{n}}} + \frac{t_{\bar{n}}\delta^2}{2},$$

we obtain with Lemma 3.5 the estimate

$$\frac{(\rho^2-1)t_{\bar{n}}\delta^2}{2} \le \inf_{p\in H}\left[J^*(K^*p) - \langle p, g\rangle + J(u^\dagger) + \frac{\|p\|^2}{2t_{\bar{n}}}\right] \le \frac{t_{\bar{n}}}{2}\Psi\left(\frac{2}{t_{\bar{n}}}\right).$$

This proves that $(\rho^2-1)\delta^2 \le \Psi(2/t_{\bar{n}})$. Now the assertion follows by applying the monotonically increasing function Ψ^{-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 Φ in Assumption 4.1 to an equivalent condition on the function $\Psi = (\Phi^{-1})^*$.

Lemma 4.5. Let Φ be an index function and Ψ the Fenchel conjugate of Φ^{-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-increasing. Because Ψ is convex and continuous, this condition is satisfied, if and only if $H'(t) \le 0$ for every $t > 0$ for which $\Psi'(t)$ exists. Now,

$$H'(t) = \frac{\Psi'(t)}{t^2} - \frac{2\Psi(t)}{t^3} = \frac{1}{t^3}\big(t\Psi'(t) - 2\Psi(t)\big),$$

and therefore $H'(t) \le 0$ if and only if $t\Psi'(t) - 2\Psi(t) \le 0$. Now recall that Ψ is the Fenchel conjugate of Φ^{-1} and therefore $t\Psi'(t) = \Psi(t) + \Phi^{-1}\big(\Psi'(t)\big)$. Thus $H'(t) \le 0$, if and only if $\Phi^{-1}\big(\Psi'(t)\big) - \Psi(t) \le 0$.

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

$$\tilde{H}'(s) = \frac{2s}{\Phi^{-1}(s)} - \frac{s^2\Phi^{-1\prime}(s)}{\Phi^{-1}(s)^2} = \frac{s\big(2\Phi^{-1}(s) - s\Phi^{-1\prime}(s)\big)}{\Phi^{-1}(s)^2} \le 0.$$

Because of the equality $s{\Phi^{-1}}'(s) = \Phi^{-1}(s) + \Psi\big({\Phi^{-1}}'(s)\big)$, this is the case, if and only if $\Phi^{-1}(s) - \Psi\big({\Phi^{-1}}'(s)\big) \le 0$. 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 Φ^{-1} and Ψ are conjugate. □

Proof of Theorem 4.2. Throughout the proof we use the abbreviation $n = n^*(\delta)$. First observe that $K^* p_n^\delta \in \partial J(u_n^\delta)$ and thus $J(u_n^\delta) - J(u^\dagger) \le \langle p_n^\delta, K u_n^\delta - g\rangle$. From the discrepancy rule (22) it follows that

$$\|K u_n^\delta - g\| \le \|K u_n^\delta - g^\delta\| + \delta \le (\rho+1)\delta,$$

and hence the variational inequality (9) implies

$$D(u_n^\delta, u^\dagger) \le \|p_n^\delta\|\,(\rho+1)\delta + \Phi((\rho+1)^2\delta^2). \tag{23}$$

As in the proof of Lemma 3.4 we observe that for all s, $r \ge 0$ one has $s\Phi(r) \le \Psi(s) + r$. 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

$$\Phi\big((\rho+1)^2\delta^2\big) \le \frac{(\rho^2-1)\delta^2}{\Psi^{-1}((\rho^2-1)\delta^2)} + \frac{(\rho+1)^2\delta^2}{\Psi^{-1}((\rho^2-1)\delta^2)} = \frac{2\rho(\rho+1)\delta^2}{\Psi^{-1}((\rho^2-1)\delta^2)},$$

which yields an estimate for the second term in (23). For estimating the first term, we note that Corollary 3.6 implies the estimate

$$\|p_n^\delta\| \le \tilde{C} t_n \left(\Psi\left(\frac{2}{t_n}\right) + \delta^2\right)^{1/2} \le \tilde{C} t_n \Psi\left(\frac{2}{t_n}\right)^{1/2} + \tilde{C} t_n \delta$$

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 Ψ,

$$\begin{aligned}
\|p_n^\delta\| &\le \tilde{C}\left(\frac{2}{\Psi^{-1}((\rho^2-1)\delta^2)} + \tau_n\right)\Psi\left(\frac{2\Psi^{-1}\big((\rho^2-1)\delta^2\big)}{2 + \tau_n \Psi^{-1}\big((\rho^2-1)\delta^2\big)}\right)^{1/2} \\
&\qquad + \frac{2\tilde{C}\delta}{\Psi^{-1}\big((\rho^2-1)\delta^2\big)} + \tilde{C}\tau_n\delta \\
&\le \frac{2\tilde{C}(\rho+1)\delta}{\Psi^{-1}\big((\rho^2-1)\delta^2\big)} + \tilde{C}\tau_n(\rho+1)\delta.
\end{aligned}$$

Consequently we have

$$\begin{aligned}
D(u_n^\delta, u^\dagger) &\le \frac{2\tilde{C}(\rho+1)^2\delta^2}{\Psi^{-1}\big((\rho^2-1)\delta^2\big)} + \tilde{C}(\rho+1)^2\tau_n\delta^2 + \frac{2\rho(\rho+1)\delta^2}{\Psi^{-1}\big((\rho^2-1)\delta^2\big)} \\
&\le \frac{2(\tilde{C}+1)(\rho+1)^2\delta^2}{\Psi^{-1}\big((\rho^2-1)\delta^2\big)} + \tilde{C}(\rho+1)^2\delta^2 \sup_k \tau_k,
\end{aligned}$$

which proves the assertion with $C := 2(\tilde{C}+1)$. □

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

$$\left\langle \xi^\dagger, u^\dagger - u \right\rangle_{X^*,X} = \left\langle p^\dagger, g - Ku \right\rangle \leq \left\| p^\dagger \right\| \|Ku - g\| .$$

Setting $\beta = 1$ and $\Phi(t) = \left\| p^\dagger \right\| t^{1/2}$ 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 a $p^\dagger \in V$ such that $K^*p^\dagger \in \partial J(u^\dagger)$. However, if (11) is replaced by the stronger condition

$$\left\langle \xi^\dagger, u^\dagger - u \right\rangle_{X^*,X} \leq (1-\beta) D_J(u, u^\dagger) + \gamma \|Ku - g\| , \tag{24}$$

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 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 $u = u^\dagger - tv$ in (24) for some $t > 0$. Then,

$$\left\langle \xi^\dagger, tv \right\rangle_{X^*,X} \leq (1-\beta) D_J(u^\dagger - tv, u^\dagger) + \gamma \|tKv\| .$$

Since the mapping $w \mapsto J'(u^\dagger)(w)$ is positively one-homogeneous, this implies that

$$\left\langle \xi^\dagger, v \right\rangle_{X^*,X} \leq (1-\beta) \left(\frac{J(u^\dagger - tv) - J(u^\dagger)}{t} - J'(u^\dagger)(-v) \right) + \gamma \|Kv\| ,$$

for all $v \in X$ and $t > 0$. Letting $t \to 0^+$ this shows that $\left\langle \xi^\dagger, v \right\rangle_{X^*,X} \leq \gamma \|Kv\|$ for all $v \in X$ and hence $K^*p^\dagger = \xi^\dagger$ for some $p^\dagger \in H$ according to (Scherzer et al. 2009, Lem. 8.21).

In the particular case where the mapping J is Gâteaux differentiable at $u^\dagger$, the subdifferential $\partial J(u^\dagger)$ contains a single element $\xi^\dagger$, which coincides with the directional derivative, that is, $\left\langle \xi^\dagger, v \right\rangle = J'(u^\dagger)(v)$ 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 Ψ of Φ^{-1} reads as $\Psi(t) = \gamma/(2\sqrt{2})t^2$. Hence it follows from Theorem 3.3 that there exists a constant $C > 0$ such that

$$D_J^{K^*p^\dagger}(u_n^\delta, u^\dagger) \leq C\delta$$

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 & Scherzer 2010)). We note that the results in (Frick & Scherzer 2010) are slightly stronger, as they give δ-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

$$D_J^{K^*p^\dagger}(u^\delta_{n^*(\delta)}, u^\dagger) \leq C\sqrt{\frac{(\rho+1)^3}{\rho-1}}\delta + C(\rho+1)^2\delta^2 \sup_{k\in\mathbb{N}} \tau_k.$$

This coincides with the results in (Frick et al. 2011, Thm. 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., $L = \mathrm{Id}$ and thus $J(u) = \frac{1}{2}\|u\|^2$. Then, a solution $u^\dagger$ of (1) is said to satisfy a Hölder condition with exponent $0 \leq \nu < \frac{1}{2}$, if $(K^*K)^\nu p^\dagger = u^\dagger = \partial J(u^\dagger)$. If $u^\dagger$ satisfies a Hölder condition with exponent ν, then (11) holds with $D_J^{u^\dagger}(u, u^\dagger) = \frac{1}{2}\|u - u^\dagger\|^2$ and $\Phi(s) \asymp s^{\frac{2\nu}{1+2\nu}}$. To see this, observe that the interpolation inequality (cf. (Engl et al. 1996, p.47)) implies

$$\begin{aligned}
\langle u^\dagger, u^\dagger - u\rangle &\leq \|p^\dagger\| \, \|(K^*K)^\nu (u^\dagger - u)\| \\
&\leq \|p^\dagger\| \, \left\|(K^*K)^{\frac{1}{2}}(u^\dagger - u)\right\|^{2\nu} \|u^\dagger - u\|^{1-2\nu} \\
&= 2^{\frac{1}{2}-\nu} \|p^\dagger\| \left(\|Ku - g\|^2\right)^\nu D_J^{u^\dagger}(u, u^\dagger)^{\frac{1-2\nu}{2}}.
\end{aligned}$$

Using Young's inequality $ab \leq a^p/p + b^q/q$ with $q = 2/(1-2\nu)$ and $p = 2/(1+2\nu)$ shows for all $\eta > 0$

$$\begin{aligned}
\left(\|Ku - g\|^2\right)^\nu D_J^{u^\dagger}(u, u^\dagger)^{\frac{1-2\nu}{2}} &= \frac{1}{\eta}\left(\|Ku - g\|^2\right)^\nu \eta D_J^{u^\dagger}(u, u^\dagger)^{\frac{1-2\nu}{2}} \\
&= \frac{1+2\nu}{2\eta^{\frac{2}{(1+2\nu)}}}\left(\|Ku - g\|^2\right)^{\frac{2\nu}{1+2\nu}} + \frac{\eta^{\frac{2}{(1-2\nu)}}(1-2\nu)}{2} D_J^{u^\dagger}(u, u^\dagger).
\end{aligned}$$

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

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

$$\langle Lu^\dagger, Lu^\dagger - Lu\rangle \leq (1-\gamma)\|Lu - Lu^\dagger\|^2 + \Phi(\|u - u^\dagger\|^2) \qquad (25)$$

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

$$\begin{aligned}\left\langle Lu^\dagger, Lu^\dagger - Lu \right\rangle &= \left\langle \omega^\dagger, \hat{L}^\nu (Lu^\dagger - Lu) \right\rangle \\ &\le \left\|\omega^\dagger\right\| \left\| \hat{L}^{\frac{1}{2}} (Lu^\dagger - Lu) \right\|^{2\nu} \left\| Lu^\dagger - Lu \right\|^{1-2\nu} \\ &\le \left\|\omega^\dagger\right\| \left\| L\tilde{L}^{\frac{1}{2}} \right\|^{2\nu} \left\| u^\dagger - u \right\|^{2\nu} \left\| Lu^\dagger - Lu \right\|^{1-2\nu} .\end{aligned}$$

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

Now let X again be a general Banach space and $J \colon X \to \overline{\mathbb{R}}$ be convex such that Assumptions 2.1 are satisfied. As revealed by the calculations above, the variational inequality (9) with $\Phi(s) \asymp s^{\frac{2\nu}{1+2\nu}}$ can be seen as a generalized Hölder condition. Note, that in this case the Legendre conjugate Ψ of Φ^{-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

$$D(u_n^\delta, u^\dagger) \le C \delta^{\frac{4\nu}{1+2\nu}}$$

if $t_n \asymp \delta^{\frac{-2}{1+2\nu}}$. Morozov's discrepancy principle (22) then shows that

$$D(u_{n^*(\delta)}^\delta, u^\dagger) \le C \left(\frac{(\rho+1)^{1+4\nu}}{\rho - 1} \right)^{\frac{1}{1+2\nu}} \delta^{\frac{4\nu}{1+2\nu}} + C(\rho+1)^2 \delta^2 \sup_{k \in \mathbb{N}} \tau_k .$$

These results coincide with the lower order rates for the iterated Tikhonov method (Hanke & 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 $1 \le q < 2$ (see (Daubechies et al. 2004)). In (Grasmair et al. 2008), it has been shown that, for Tikhonov regularization, this setting allows the derivation of convergence rates of order $\mathcal{O}(\delta^q)$ with respect to the norm, if $u^\dagger$ satisfies the standard source condition $K^* p^\dagger \in \partial J(u^\dagger)$ for some $p^\dagger \in H$, and, additionally, a *restricted injectivity condition* holds. In the following, we will generalize these results to the Augmented Lagrangian Method and source conditions of Hölder type.

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

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

$$C_1 \left\| u^\dagger - u \right\|^q \le C_2 \left\| Ku - g \right\|^q + J(u) - J(u^\dagger) - \left\langle \xi^\dagger, u - u^\dagger \right\rangle$$

for all $u \in X$. Applying the interpolation inequality to $\left\langle \xi^\dagger, u - u^\dagger \right\rangle$, we obtain, similarly as in Section 5.2, the estimate

$$C_1 \left\| u^\dagger - u \right\|^q \le C_2 \left\| Ku - g \right\|^q + J(u) - J(u^\dagger) + \left\| p^\dagger \right\| \left\| Ku - g \right\|^{2\nu} \left\| u^\dagger - u \right\|^{1-2\nu} .$$

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

$$\left[C_1 - \left\|p^\dagger\right\| \frac{1-2\nu}{q}\eta^{\frac{q}{1-2\nu}}\right] \left\|u^\dagger - u\right\|^q \le C_2 \left\|Ku-g\right\|^q + J(u) - J(u^\dagger) + \left\|p^\dagger\right\| \frac{q-1+2\nu}{q}\eta^{\frac{q}{q-1+2\nu}} \left\|Ku-g\right\|^{\frac{2\nu q}{q-1+2\nu}} .$$

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

$$\Phi(s) = C_2 s^{\frac{q}{2}} + \left\|p^\dagger\right\| \frac{q-1+2\nu}{q}\eta^{\frac{q}{q-1+2\nu}} s^{\frac{2\nu q}{q-1+2\nu}},$$

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

$$\left\|u_n^\delta - u^\dagger\right\| \le C\delta^{\frac{2\nu}{q-1+2\nu}}$$

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

Remark 5.1. In (Grasmair et al. 2011), it has been shown for Tikhonov regularization with $J(u) = \sum_i |\langle \phi_i, u\rangle|$, 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 results above imply that, in the case $q = 1$, the Hölder type source condition $(K^*K)^\nu p^\dagger \in \partial J(u^\dagger)$ in fact already implies the standard source condition $K^*\tilde{p}^\dagger \in \partial J(u^\dagger)$ for some different source element $\tilde{p}^\dagger$.

Acknowledgments

The second author would like to thank Axel Munk and the staff of the Institute for Mathematical Stochastics at the 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* Vol. 7 of *Studies in Mathematics and its Applications* North-Holland Publishing Co. Amsterdam.

Bachmayr M & Burger M 2009 *Inverse Problems* **25**(10), 105004, 26.

Boţ R I & Hofmann B 2010 *J. Integral Equations Appl.* **22**(3), 369–392.

Brill M & Schock E 1987 *in* 'Model optimization in exploration geophysics (Berlin, 1986)' Vol. 1 of *Theory Practice Appl. Geophys.* Vieweg Braunschweig pp. 13–37.

Burger M & Osher S 2004 *Inverse Problems* **20**(5), 1411–1421.

Burger M, Resmerita E & He L 2007 *Computing* **81**(2-3), 109–135.

Daubechies I, Defrise M & De Mol C 2004 *Comm. Pure Appl. Math.* **57**(11), 1413–1457.

Ekeland I & Temam R 1976 *Convex analysis and variational problems* Vol. 1 of *Studies in Mathematics and its Applications* North-Holland Publishing Co. Amsterdam-Oxford.

Engl H, Hanke M & Neubauer A 1996 *Regularization of inverse problems* Mathematics and its Applications Kluwer Academic Publishers Group Dordrecht.

Fortin M & Glowinski R 1983 *Augmented Lagrangian methods* Vol. 15 of *Studies in Mathematics and its Applications* North-Holland Publishing Co. Amsterdam. Applications to the numerical solution of boundary value problems, Translated from the French by B. Hunt and D. C. Spicer.

Frick K, Lorenz D A & Resmerita E 2011 *Multiscale Modeling & Simulation* **9**(4), 1528–1548.

Frick K & Scherzer O 2010 *J. Integral Equations Appl.* **22**(2), 217–257.

Grasmair M 2010 *Inverse Problems* **26**(11), 115014.

Grasmair M, Haltmeier M & Scherzer O 2008 *Inverse Problems* **24**(5), 055020, 13.

Grasmair M, Haltmeier M & Scherzer O 2011 *Comm. Pure Appl. Math.* **64**(2), 161–182.

Groetsch C W 2007 *Stable approximate evaluation of unbounded operators* number 1894 *in* 'Lecture Notes in Mathematics' Springer-Verlag Berlin.

Groetsch C W & Scherzer O 2000 *Math. Methods Appl. Sci.* **23**(15), 1287–1300.

Güler O 1991 *SIAM J. Control Optim.* **29**(2), 403–419.

Hanke M & Groetsch C W 1998 *J. Optim. Theory Appl.* **98**(1), 37–53.

Hestenes M R 1969 *J. Optimization Theory Appl.* **4**, 303–320.

Hofmann B, Kaltenbacher B, Pöschl C & Scherzer O 2007 *Inverse Problems* **23**(3), 987–1010.

Ito K & Kunisch K 2008 *Lagrange multiplier approach to variational problems and applications* Vol. 15 of *Advances in Design and Control* Society for Industrial and Applied Mathematics (SIAM) Philadelphia, PA.

Jung M, Resmerita E & Vese L 2011 *J. Math. Imag. Vis.* pp. 1–22.

Krasnosel′skiĭ M A 1960 *Uspehi Mat. Nauk* **15**(3 (93)), 161–165.

Krjanev A V 1973 *Dokl. Akad. Nauk SSSR* **210**, 20–22.

Lardy L J 1975 *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8)* **58**(2), 152–157.

Martinet B 1970 *Rev. Française Informat. Recherche Opérationnelle* **4**(Ser. R-3), 154–158.

Morozov V A 1967 *Doklady Akademii Nauk SSSR* **175**(6), 1225–1228.

Osher S, Burger M, Goldfarb D, Xu J & Yin W 2005 *Multiscale Model. Simul.* **4**(2), 460–489 (electronic).

Powell M J D 1969 *in* 'Optimization (Sympos., Univ. Keele, Keele, 1968)' Academic Press London pp. 283–298.

Rockafellar R T 1974 *SIAM J. Control* **12**, 268–285.

Scherzer O, Grasmair M, Grossauer H, Haltmeier M & Lenzen F 2009 *Variational methods in imaging* Vol. 167 of *Applied Mathematical Sciences* Springer New York.