Research Network FWF S105

Photoacoustic Imaging in
Medicine and Biology

http://pai.uibk.ac.at

Regularization of Ill-Posed Linear Equations by the Non-Stationary Augmented Lagrangian Method

Klaus Frick and Otmar Scherzer

September 2009

PAI Report No. 20

Der Wissenschaftsfonds.

REGULARIZATION OF ILL-POSED LINEAR EQUATIONS BY THE NON-STATIONARY AUGMENTED LAGRANGIAN METHOD

KLAUS FRICK AND OTMAR SCHERZER

Abstract. In this paper, we make a convergence rates analysis of the non-stationary Augmented Lagrangian Method for the solution of linear inverse problems. The motivation for the analysis is the fact that the Tikhonov-Morozov Method is a special instance of the Augmented Lagrangian Method. In turn, the later is also equivalent to iterative Bregman distance regularization, which received much attention in the imaging literature recently.

We base the analysis of the Augmented Lagrangian Method on convex duality arguments. Thereby, we can reprove some of the convergence (rates) results for the Tikhonov Morozov Method. In addition, by the novel analysis we can prove novel convergence and convergence rates results for the dual variables of the Augmented Lagrangian methods. Reinterpretation of the dual variables for the Tikhonov-Morozov Method gives some new convergence rates results for the linear functionals of the regularized solutions. As a benchmark for achievable convergence rates of the Augmented Lagrangian Method in the general convex context we use the results on evaluation of unbounded operators of Groetsch [13], which is a special instance of the Tikhonov-Morozov method. In addition we derive the flow, which interpolates the iterates of the Augmented Lagrangian Method and show the relation to Showalter's method.

1. Introduction

In this paper, we are concerned with solving constrained optimization problems

$$J(u) \to \min \qquad \text{subject to} \qquad Ku = g\,, \tag{1}$$

where J is a convex functional and $K : H_1 \to H_2$ is a linear bounded operator between Hilbert space H_1 and H_2. Minimizers of (1) are called J-minimizing solutions of the equation

$$Ku = g. \tag{2}$$

Our main interests are ill–posed equations, that is, when the solution of (2) does not depend continuously on the data g. Our analysis takes into account data perturbations in g, which are denoted by g^δ, for which we assume that we have the additional information that

$$\left\|g^\delta - g\right\| \leq \delta. \tag{3}$$

The prime application of the *Augmented Lagrangian Method* (the ALM) is to solve constrained optimization problems of the form (1) and reads as follows:

Algorithm 1.1 (the ALM)**.** Let $p_0^\delta \in H_2$ and choose a sequence $\{\tau_n\}_{n\in\mathbb{N}}$ of positive parameters. For $n = 1, 2, \dots$ compute

$$u_n^\delta \in \operatorname*{argmin}_{u\in H_1} \left(\frac{\tau_n}{2} \left\|Ku - g^\delta\right\|^2 + J(u) - \left\langle p_{n-1}^\delta, Ku - g^\delta \right\rangle \right) \qquad \text{and} \tag{4a}$$

$$p_n^\delta = p_{n-1}^\delta + \tau_n (g^\delta - K u_n^\delta). \tag{4b}$$

Historically, the ALM dates back to Hestenes [16] and Powell [25] (there called *method of multipliers*). For background references on the ALM we refer to Fortin & Glowinski [10] and the recent book by Ito & Kunisch [20]. In the context of imaging and total variation regularization the ALM has been considered for instance in [19].

The *Tikhonov-Morozov* method is another examples of a regularization method for solving constrained minimization problems. Our subjective opinion is that the Tikhonov-Morozov theory has *not* been considered in the same general setting as the ALM. However the theoretical results, especially convergence rates, seem to go much beyond the theory of the ALM. We will show this by a comparison of the results of the respective fields.

We follow the relevant literature of Tikhonov-Morozov regularization and work with a linear and closed operator $L : D(L) \subset H_1 \to H$, whose domain of definition $D(L)$ is a dense subset of H_1. For the comparison with the ALM we use the convex functional

$$J(u) = \frac{1}{2} \|Lu\|^2 \,, \tag{5}$$

and then the Tikhonov-Morozov method consists in choosing $u_0^\delta \in H_1$ and by iteratively calculating

$$\boxed{u_n^\delta := \operatorname*{argmin}_{u \in H_1} \left(\frac{\tau_n}{2} \left\| Ku - g^\delta \right\|^2 + J(u - u_{n-1}^\delta) \right).} \tag{6}$$

It is common to differ between *stationary* and *non-stationary* methods, depending on whether the parameters τ_n are chosen constant or variable.

The present paper shows that Tikhonov-Morozov regularization and the ALM, with convex penalization functional J from (5), are equivalent. Recently, there have been several publications revealing the equivalence relation between the ALM and iterative *Bregman-Distance Regularization* (see Setzer [29] and the first authors's thesis [11]), which consists in iterative calculation of

$$\boxed{u_n^\delta := \operatorname*{argmin}_{u \in H_1} \left(\frac{\tau_n}{2} \left\| Ku - g^\delta \right\|^2 + D_J^{K^* p_{n-1}^\delta}(u, u_{n-1}^\delta) \right) ,}$$

where $D_J^\xi(u,v)$ denotes the Bregman-distance between u and v with respect to ξ (cf. Section 2). Bregman Distance Regularization has been suggested in [24] and analyzed from a regularization theoretic point of view, with the background of total variation minimization in [6]. Note, however, that general results for the ALM apply also to Bregman Distance Regularization, due to cited equivalence relations. As a further consequence of the above discussion Tikhonov-Morozov regularization is equivalent to Bregman distance regularization when the penalization functional J from (5) is used. Therefore, all results in this paper, which we derive for the ALM, are equally valid for Bregman Distance Regularization, and in particular for the Tikhonov-Morozov method.

Moreover, we show that by using Rockafeller's duality concept [27] the dual iterates $\{p_n^\delta\}_{n\in\mathbb{N}}$ of the ALM can be rewritten as minimizers of a generalized iterated Tikhonov-Morozov regularization method, the so called *proximal point method*. By generalized we mean that the fit-to-data term is a general convex functional $G(\cdot, g^\delta)$ and not the square of a norm. In general the convergence rate analysis of regularization methods with general fit-to-data terms is by far from being as complete in comparison with quadratic regularization. The outlined relations, however, given an indication of a convergence rates analysis of the ALM based on dual variables. In particular, as we show, making a convergence rates analysis for the dual iterates allows for deriving convergence rates for the primal iterates. This is the main contribution of this paper.

To summarize, we show below that the ALM is equivalent to a generalized iterative Tikhonov-Morozov method for dual variables.

An important application of Tikhonov-Morozov regularization is the evaluation of an unbounded operator L, which is a standard example of an ill–posed problem. In this case we have the particular situation of above with $K = \mathrm{Id}$. A typical example of an unbounded operator L is the Moore-Penrose inverse (see [22]) of a compact linear operator. For evaluation of unbounded operators, Morozov [21] proposed a regularization method consisting in calculating Lu_α^δ, where

$$\boxed{u_\alpha^\delta := \operatorname{argmin} \left(\left\| u - u^\delta \right\|^2 + \tau^{-1} \|Lu\|^2 \right) ,} \tag{7}$$

for some $\tau > 0$. This is equivalent to computing $u_\alpha^\delta = (\tau + L^*L)^{-1} u^\delta$. The major player in the field of analysis of regularization methods for evaluation of unbounded operators is C. W. Groetsch (see his monograph [13]): For instance he proved optimal convergence rates up to maximal order $O(\delta^{2/3})$. Faster convergence rates are possible for the *iterative Tikhonov-Morozov* method, which uses for approximation of Lu the evaluations of the iterates u_n^δ, which are then defined as the minimizers of the functional

$$\boxed{u \to \left(\left\| u - u^\delta \right\|^2 + \tau_n^{-1} \left\| Lu - Lu_{n-1}^\delta \right\|^2 \right).} \tag{8}$$

Here, in contrast to the standard Morozov regularization, it has been shown that with an appropriate choice of the regularization parameters, $\{\tau_n\}_{n\in\mathbb{N}}$ and the stopping iteration n_*, both depending on δ, convergence rates up to order δ are possible [13].

We believe that many convergence and stability results known for Tikhonov-Morozov regularization are still open for the ALM - that is, in the general convex setting. This paper makes the attempt to point out some of the open issues which can be further considered when generalizing variational regularization theory for unbounded operators and for the Tikhonov-Morozov method. In this sense, the work of Groetsch [13] serves as a benchmark on achievable results in the general setting. A novel facet of the convergence analysis is that we add (weak) convergence and rates results for the dual variables of the ALM. Currently, in fact, convergence rates results for the Tikhonov-Morozov method are just expressed with respect to the residuals and the iterates. In the context of Tikhonov-Morozov method, the errors of the dual variables can be expressed as functionals of the residuals, for which convergence rates then follow from the theory of the ALM. It will become transparent that convergence rates of residuals, iterates, and dual variables are strongly coupled.

Finally, we investigate asymptotic methods, which interpolate the iterates of the ALM. Since the ALM and Bregman distance regularization are equivalent, we call the resulting continuous regularization method *Bregman distance flow* (see for instance [4] for examples of such flows). With the functional (5) this flow method resembles the *Showalter method*, as a method, which interpolates the Tikhonov-Morozov method. Again, we use convex duality arguments combined with standard results from semi-group theory for proving existence of solutions of flows.

2. Basic Definitions

The aim of this section is to summarize the basic definitions and assumptions needed to perform a convergence analysis of the ALM. We use the following basic assumptions and notions from convex analysis:

Assumption 2.1. (i) H_1, H_2, and H denote non-empty Hilbert spaces. The norms on H_1, H_2, and H, respectively are not further specified, and will be always denoted by $\|\cdot\|$, since the meaning is clear from the context.

(ii) Let $J : H_1 \to \overline{\mathbb{R}}$ be a convex functional from H_1 into the extended reals $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$. The domain of J is defined by

$$D(J) = \{u \in H_1 : J(u) \neq \infty\} \ .$$

J is called *proper* if $D(J) \neq \emptyset$ and $J(u) > -\infty$ for all $u \in H_1$. Throughout this paper J denotes a convex, proper and lower semi-continuous (l.s.c.) functional.

(iii) $K : H_1 \to H_2$ is a linear and bounded operator.

In the course of this paper we will frequently make use of tools from convex analysis. For a standard reference we refer to Ekeland & Temam [8].

- The subdifferential (or generalized derivative) $\partial J(u)$ of J at u is the set of all elements $p \in H_1$ satisfying

$$J(v) - J(u) - \langle p, v - u \rangle \geq 0.$$

When the subgradient consists of a single element, here and in the following, we identify the subgradient with the element. The *domain* $D(\partial J)$ of the subgradient consists of all $u \in H_1$ for which $\partial J(u) \neq \emptyset$. Finally, we define the *graph* of ∂J as

$$\mathrm{Gr}(\partial J) := \{(u,p) \in H_1 \times H_1 \ : \ p \in \partial J(u)\} \, .$$

According to [8, Chap.1 Cor 5.1], the set $\mathrm{Gr}(\partial J)$ is sequentially closed w.r.t. strong-weak topology on $H_1 \times H_1$. That is, if the sequence $\{(u_n, v_n)\}_{n\in\mathbb{N}}$ of elements in $\mathrm{Gr}(\partial J)$ satisfies that u_n converges weakly to u and v_n converges strongly to v, then $(u,v) \in \mathrm{Gr}(\partial J)$.

- The functional $J^* : H_1 \to \overline{\mathbb{R}}$ denotes the Legendre-Fenchel transform (or the dual functional) of J, which is defined by
$$J^*(v) := \sup_{u\in H_1} (\langle v, u\rangle - J(u)). \tag{9}$$
From its definition it becomes clear, that J^* is the pointwise supremum of affine functions and thus, according to [8, Chap I, Prop. 3.1], convex, l.s.c. and proper. Moreover, one has [8, Chap I, Cor. 5.2.]
$$p \in \partial J(u) \Leftrightarrow u \in \partial J^*(p). \tag{10}$$
- Typically, convergence of the ALM is proven with respect to the *Bregman-distance*. For $u \in D(\partial J)$ the Bregman distance of J between u and v w.r.t to $\xi \in \partial J(u)$ is defined by
$$D_J^\xi(v,u) = J(v) - J(u) - \langle \xi, v-u\rangle\,.$$

Next, we introduce different classes of solutions for Equation (2) discussed in this paper.

Definition 2.2. (i) An element $u \in H_1$ satisfying (2) is called a *solution* of (2).
(ii) Let $u \in D(J)$ be a solution of $g = Ku$. Then g is called *attainable.*
(iii) An element $u \in D(J)$ is called *J-minimizing solution* of (2), if u solves (1).
(iv) Let $g \in H_2$ be attainable. An element $p \in H_2$ is called a *source element* if there exists a J-minimizing solution u of (2) such that
$$K^*p \in \partial J(u). \tag{11}$$
Then, we say hat u *satisfies the source condition* (11).

In general, Assumption 2.1 is not enough to guarantee existence of J-minimizing solutions or the well-posedness of the ALM. For that, one needs a coercivity condition, like the following:

Assumption 2.3. For each $c \in \mathbb{R}$, the sub-level sets of the functional
$$u \to \|Ku\|^2 + J(u)$$
are sequentially pre-compact with respect to the weak topology on H_1. That is, for every $c \in \mathbb{R}$, every sequence $\{u_n\}_{n\in\mathbb{N}}$ contained in the sub-level set
$$\Lambda(c) = \left\{u \in H_1 \;:\; \|Ku\|^2 + J(u) \le c\right\}$$
has a weak convergent subsequence in H_1.

In the remainder of this section we discuss the notions introduced above for the particular example, when J is chosen as in (5), i.e. when the ALM reduces to the Tikhonov-Morozov method.

As it is common in the theory of the Tikhonov-Morozov method [21] and the evaluation of unbounded operators [13], let $L : D(L) \subset H_1 \to H$ be a linear and closed operator defined on the dense subset $D(L) \subset H_1 \neq \emptyset$. Closed means that the graph of L,
$$\mathrm{Gr}(L) = \{(u,v) \in H_1 \times H \;:\; L(u) = v\}$$
is sequentially closed in $H_1 \times H$. Since $D(L)$ is assumed to be dense, there exists an *adjoint* operator [34, Chap. VII.2]
$$L^* : D(L^*) \subset H \to H_1$$
with domain
$$D(L^*) := \{v \in H \;:\; u \mapsto \langle Lu, v\rangle \text{ is continuous}\}$$
satisfying
$$\langle Lu, v\rangle = \langle x, L^*v\rangle \qquad \text{for all } u \in D(L),\, v \in D(L^*).$$

In this context, the precise meaning of $J : H_1 \to \overline{\mathbb{R}}$ as in (5) is as follows:

$$J(u) = \begin{cases} \frac{1}{2}\|Lu\|^2 & \text{if } u \in D(L) \\ +\infty & \text{else.} \end{cases} \tag{12}$$

In the following lemma we give a characterization of the subgradient, the Bregman distance, and the domain $D(\partial J)$ of J as defined in (12).

Lemma 2.4. The functional J is proper, convex, and l.s.c. Moreover, $D(\partial J) = D(L^*L)$ and the subgradient is given by

$$\partial J(u) = \begin{cases} L^*Lu & \text{if } u \in D(L^*L) \\ \emptyset & \text{else.} \end{cases}$$

Proof. Since L is densely defined on a non-empty Hilbert space, J is proper. Moreover, convexity follows from the linearity of L and the convexity of $\frac{1}{2}\|\cdot\|^2$.

We prove lower semi-continuity: Assume that $\{u_n^\delta\}_{n\in\mathbb{N}}$ is a convergent sequence with limit u in H_1. If $\{Lu_n^\delta\}_{n\in\mathbb{N}}$ converges in H to an element v, then from the closedness of L it follows that $Lu = v$. Therefore, $\|Lu\|^2 = \lim \|Lu_n^\delta\| = \liminf_{n\to\infty} \|Lu_n^\delta\|$.

If Lu_n^δ does not converge, then we differ between the case that there exists a subsequence of $\left\{u_{\rho(n)}^\delta\right\}_{n\in\mathbb{N}}$ such that $\liminf_{n\to\infty} J(u_{\rho(n)}^\delta) = +\infty$ or $\{u_n^\delta\}_{n\in\mathbb{N}}$ is bounded. In the first case it is obvious that $J(v) \le \liminf_{n\to\infty} J(u_{\rho(n)}^\delta) = +\infty$ and nothing remains to be shown.

In the second case, we can select a sub-sequence $\left\{Lu_{\rho(n)}^\delta\right\}_{n\in\mathbb{N}}$ which is weakly convergent to some $v \in H_2$. Since $\mathrm{Gr}(L) \subset H_1 \times H$ is closed and convex it is weakly closed (cf. [34, Thm. III.3.8]) and thus $u_{\rho(n)} \to u$ implies $Lu = v$. Weak lower semi-continuity of the norm eventually gives

$$J(u) = \frac{1}{2}\|Lu\|^2 \le \liminf_{n\to\infty} \frac{1}{2}\left\|Lu_{\rho(n)}\right\|^2.$$

It remains to show that $D(\partial J) = D(L^*L)$ and $\partial J(u) = L^*Lu$ for $u \in D(L^*L)$. By verification it follows that $L^*Lu \in \partial J(u)$, whenever $u \in D(L^*L)$. Next, we prove that the operator $\mathrm{Gr}(L^*L)$ is maximal monotone. That is, $\mathrm{Gr}(L^*L)$ is not properly contained in any monotone set in $H_1 \times H_1$. The elementary inequality

$$\langle L^*Lu_1 - L^*Lu_2, u_1 - u_2\rangle = \|Lu_1 - Lu_2\|^2 \ge 0$$

for all $u_i \in D(L^*L)$ $(i = 1, 2)$, shows that $\mathrm{Gr}(L^*L)$ is a monotone subset of $H_1 \times H_1$. Since $D(L)$ is dense by assumption, L^*L is densely defined and self-adjoint and therefore closed in $H_1 \times H_1$ ([34, Cor. VII 2.13]). This, however, is already sufficient for $\mathrm{Gr}(L^*L)$ to be maximal monotone (see e.g. [18, Chap.3 Thm.1.45]) The subgradient of a convex and l.s.c. functional is maximal monotone Since ∂J is (maximal) monotone and due to the fact that

$$L^*Lu \subset \partial J(u)$$

this shows $\{L^*Lu\} = \partial J(u)$ for all $u \in D(L^*L)$. □

The next remark concerns the Bregman distance of J as in (12) as well as the interpretations of the notions of Definition 2.2:

Remark 2.5. Let J be as in (12).

(i) For $v \in D(L)$ and $u \in D(L^*L)$, we have

$$D_J^{L^*Lu}(v, u) = \frac{1}{2}\|L(u - v)\|^2.$$

(ii) An element $g \in H_2$ is attainable, if $K^{-1}(\{g\}) \cap D(L) \neq \emptyset$.

(iii) A J-minimizing solution u satisfies the source condition (11), if $u \in D(L^*L)$ and if there exists $p \in H_2$ such that

$$K^*p = L^*Lu.$$

3. Well-posedness and the Equivalence of ALM and Bregman-distance regularization

In this section we review results on well-definedness and monotonicity properties of the ALM (cf. Algorithm 1.1). Proving well-definedness of the ALM reduces to proving that there exists a minimizer in (4a). We will do this, by proving that for arbitrary $q, f \in H_2$ the functional

$$\begin{aligned} I(u) &:= \frac{\tau}{2}\|Ku - f\|^2 + J(u) - \langle q, Ku - f\rangle \\ &= \frac{\tau}{2}\left\|Ku - (f + \tau^{-1}q)\right\|^2 + J(u) - \tau^{-1}\|q\|^2 \end{aligned} \tag{13}$$

has a minimizer.

Theorem 3.1. *Let Assumptions 2.1 and 2.3 hold. Moreover, let $g \in H_2$ be attainable and $g^\delta \in H_2$. Then, there exists a J-minimizing solution of (2) and Algorithm 1.1 is well defined. Moreover,*

$$\boxed{K^* p_n^\delta \in \partial J(u_n^\delta) \quad \text{for all} \quad n = 1, 2, \dots} \tag{14}$$

Proof. Let $C \subset H_1$ be a closed and convex set such $D(J) \cap C \neq \emptyset$ and let $q, f \in H_2$ and $\tau > 0$. We prove that the functional I has a minimizer in C. Then, application with $C = H_1$, $f = g^\delta$ and $q = p_{n-1}^\delta$ gives well-posedness of the ALM and with $f = g$ and $C = K^{-1}(\{g\})$ gives existence of a J-minimizing solution.

Let $\{u_k\}_{k\in\mathbb{N}}$ be a minimizing sequence in C. Then it follows that

$$\sup \frac{\tau}{2}\left(\left\|Ku_k - (f + \tau^{-1}q)\right\|^2 + J(u_k)\right) < \infty$$

and consequently that $u_k \in \Lambda(c)$ for all $k \in \mathbb{N}$ and a suitably chosen $c \in \mathbb{R}$. Then, by Assumption 2.3, we can select a weakly convergent subsequence indexed by $\rho(n)$ and with weak limit $\hat{u}$. Since C is closed and convex it is weakly closed and therefore $\hat{u} \in C$. Moreover, weak lower semi-continuity of J implies weak lower semi-continuity of I and thus

$$I(\hat{u}) \leq \liminf_{k\to\infty} I(u_{\rho(k)}) = \inf_{u\in C} I(u).$$

Hence, $\hat{u}$ minimizes I over C.

For proving the second assertion, from (4a) we see that the optimality condition for u_n^δ is

$$K^* p_{n-1}^\delta - \tau_n K^*(Ku_n^\delta - g^\delta) \in \partial J(u_n^\delta).$$

By application of K^* to both sides of (4b) it follows that

$$K^* p_n^\delta = K^* p_{n-1}^\delta - \tau_n K^*(Ku_n^\delta - g^\delta).$$

Combination of both inclusions shows that

$$K^* p_n^\delta \in \partial J(u_n^\delta) \quad \text{for all } n \in \mathbb{N}.$$

□

Remark 3.2. As we have already used in the proof above, the minimizer of (4a) is not affected by adding constant functionals with respect to u to the objective functional. In such a way we can formulate an equivalent minimization problems to (4a) by adding the term

$$-\left(J(u_{n-1}^\delta) + \left\langle p_{n-1}^\delta, g^\delta + Ku_{n-1}^\delta\right\rangle\right).$$

The modified optimization problem then results in Bregman distance regularization

$$\boxed{\begin{aligned} u_n^\delta &\in \operatorname*{argmin}_{u\in H_1}\left(\frac{\tau_n}{2}\left\|Ku - g^\delta\right\|^2 + J(u) - J(u_{n-1}^\delta) - \left\langle K^* p_{n-1}^\delta, u - u_{n-1}^\delta\right\rangle\right) \\ &= \operatorname*{argmin}_{u\in H_1}\left(\frac{\tau_n}{2}\left\|Ku - g^\delta\right\|^2 + D_J^{K^* p_{n-1}^\delta}(u, u_{n-1}^\delta)\right), \end{aligned}} \tag{15}$$

Thus the ALM is equivalent to Bregman-distance regularization and the results for the respective other method apply.

We close this section with the basic observation that the residuals $\|Ku_n^\delta - g^\delta\|$ in the ALM are non-increasing. We also note that a uniform bound for the residuals can be given, provided that the initial multiplier p_0^δ in the ALM satisfies appropriate restrictions.

Corollary 3.3. For $p_0^\delta \in H_2$ the iterates of the ALM satisfy

$$\|Ku_n^\delta - g^\delta\| \le \|Ku_{n-1}^\delta - g^\delta\| , \quad n = 2, 3, \dots \tag{16}$$

If, in addition, p_0^δ satisfies

$$K^* p_0^\delta \in \partial J(u_0^\delta) , \tag{17}$$

then the inequality (16) also holds for $n = 1$.

Proof. From Theorem 3.1 we know that $K^* p_{n-1}^\delta \in \partial J(u_{n-1}^\delta)$ for all $n = 2, 3, \dots$. The definition of the subgradient hence gives for all $u \in U$

$$J(u) - J(u_{n-1}^\delta) + \langle K^* p_{n-1}^\delta, u_{n-1}^\delta - u \rangle \ge 0.$$

Then, by choosing $u = u_n^\delta$ we find that

$$\begin{aligned}
&\frac{\tau_n}{2} \|Ku_n^\delta - g^\delta\|^2 \\
\le& \frac{\tau_n}{2} \|Ku_n^\delta - g^\delta\|^2 + J(u_n^\delta) - J(u_{n-1}^\delta) + \langle K^* p_{n-1}^\delta, u_{n-1}^\delta - u_n^\delta \rangle \\
=& \frac{\tau_n}{2} \|Ku_n^\delta - g^\delta\|^2 + J(u_n^\delta) - \langle p_{n-1}^\delta, Ku_n^\delta - g^\delta \rangle - J(u_{n-1}^\delta) + \langle p_{n-1}^\delta, Ku_{n-1}^\delta - g^\delta \rangle \ .
\end{aligned} \tag{18}$$

From the definition of the ALM it follows that

$$\begin{aligned}
&\frac{\tau_n}{2} \|Ku_n^\delta - g^\delta\|^2 + J(u_n^\delta) - \langle p_{n-1}^\delta, Ku_n^\delta - g^\delta \rangle \\
\le& \frac{\tau_n}{2} \|Ku_{n-1}^\delta - g^\delta\|^2 + J(u_{n-1}^\delta) - \langle p_{n-1}^\delta, Ku_{n-1}^\delta - g^\delta \rangle \ .
\end{aligned}$$

Using this estimate together with (18) shows that

$$\frac{\tau_n}{2} \|Ku_n^\delta - g^\delta\|^2 \le \frac{\tau_n}{2} \|Ku_{n-1}^\delta - g^\delta\|^2 .$$

The proof of the second assertion uses the additional assumption on p_0^δ which makes the above proof applicable also in the case $n = 1$.

□

We remark that from the definition of the ALM it follows that

$$\frac{\|p_n^\delta - p_{n-1}^\delta\|}{\tau_n} = \|Ku_n^\delta - g^\delta\|$$

and therefore the scaled difference of dual variables is decreasing.

4. Duality: ALM and the Proximal Point Method

We review a duality concept due to Rockafellar [27], which characterizes the sequence $\{p_n^\delta\}_{n\in\mathbb{N}}$ in the ALM by the *proximal point method*. This will be the key to the convergence analysis of the ALM on the one hand (cf. Sections 5 and 6) and to the analysis of related evolution equations (cf. Section 7) on the other hand.

To show this relation, we introduce the *descent functional* $G : H_2 \times H_2 \to \overline{\mathbb{R}}$, defined by

$$G(p, g) := (J^* \circ K^*)(p) - \langle p, g \rangle . \tag{19}$$

The descent functional (19) exhibits the following properties:

Lemma 4.1. For $g \in H_2$ let

$$F(h, g) := \begin{cases} \inf \{J(v) \ : \ v \in H_1, \, Kv = g + h\} & \text{if } g + h \text{ is attainable} \\ +\infty & \text{else.} \end{cases}$$

Then

$$G(p, g) = F^*(\cdot, g)(p) ,$$

where $F^*(\cdot, g)$ denotes the Fenchel dual of F with respect to the first variable. In particular, if g is attainable, $G(\cdot, g)$ is bounded from below.

Proof. Let $g \in H_2$ and define $A_g \subset H_2$ be the collection of all $h \in H_2$, such that $g+h$ is attainable. Since J is assumed to be proper, we have that $A_g \neq \emptyset$ and from the definition of the Legendre-Fenchel transform in (9) and the definition of F it follows

$$\begin{aligned} F^*(\cdot, g)(p) &= \sup_{h \in H_2} \left(\langle p, h \rangle - F(h, g) \right) \\ &= \sup_{h \in A_g} \left(\langle p, h \rangle - \inf \left\{ J(v) \ : \ v \in H_1,\, Kv = g + h \right\} \right) \end{aligned}$$

The last term can be rewritten to

$$\sup_{h \in A_g} \left(\langle p, h \rangle - \inf \left\{ J(v) \ : \ v \in H_1,\, Kv = g + h \right\} \right) = \sup_{h \in A_g} \sup_{Kv = g+h} \left(\langle p, h \rangle - J(v) \right).$$

Now, taking into account that the second supremum is taken over all v that satisfy $h = Kv - g$ it follows that

$$\sup_{h \in A_g} \sup_{Kv = g+h} \left(\langle p, h \rangle - J(v) \right) = \sup_{h \in A_g} \sup_{Kv = g+h} \left(\langle K^* p, v \rangle - J(v) \right) - \langle p, g \rangle. \tag{20}$$

Therefore, eventually, we find by using again the definition of the Legendre-Fenchel transform

$$\begin{aligned} F^*(\cdot, g)(p) &= \sup_{v \in H_1} \left(\langle K^* p, v \rangle - J(v) \right) - \langle p, g \rangle \\ &= J^*(K^* p) - \langle p, g \rangle \\ &= G(p, g). \end{aligned}$$

If g is attainable, there exists $v_0 \in D(J)$ such that $Kv_0 = g$, and then it follows from (20) by estimating the supremum of the functional with the evaluation at $h = 0$ that

$$G(p, g) \geq -\inf \left\{ J(v) \ : \ v \in H_1,\, Kv = g \right\} \geq -J(v_0).$$

Since $v_0 \in D(J)$, $G(\cdot, g)$ is bounded from below. □

In the following, we derive an equivalent characterization for the dual variables $\left\{ p_n^\delta \right\}_{n \in \mathbb{N}}$ of the ALM, which is independent of the primal variables of $\left\{ u_n^\delta \right\}_{n \in \mathbb{N}}$. This observation dates back to the work of Rockafellar in [27].

Proposition 4.2. *Let $g^\delta \in H_2$. Then, for $n = 1, 2, \ldots$*

$$\boxed{p_n^\delta = \operatorname*{argmin}_{p \in H_2} \left(\frac{1}{2} \left\| p - p_{n-1}^\delta \right\|^2 + \tau_n G(p, g^\delta) \right).} \tag{21}$$

Proof. From (14) we find that $K^* p_n^\delta \in \partial J(u_n^\delta)$ for all $n = 2, 3, \ldots$ and thus it follows from (10) that

$$u_n^\delta \in \partial J^* \left(K^* p_n^\delta \right).$$

Then, from the definition of a subdifferential, it follows that for all $\xi \in H_1$

$$\left\langle u_n^\delta, \xi - K^* p_n^\delta \right\rangle + J^*(K^* p_n^\delta) \leq J^*(\xi).$$

For $p \in H_2$ and $\xi = K^* p$ we then get

$$\left\langle K u_n^\delta, p - p_n^\delta \right\rangle = \left\langle u_n^\delta, K^* p - K^* p_n^\delta \right\rangle \leq J(K^* p) - J^*(K^* p_n^\delta)$$

which is equivalent to that $K u_n^\delta \in \partial \left(J^* \circ K^* \right)(p_n^\delta)$. Taking into account the definition of p_n^δ in the ALM, it follows that

$$\tau_n^{-1} (p_{n-1}^\delta - p_n^\delta) \in \partial \left(J^* \circ K^* \right)(p_n^\delta) - g^\delta = \partial G(\cdot, g^\delta)(p_n^\delta).$$

which is equivalent to that (21) holds. □

We emphasize, that in the regularization community the determination of p_n^δ is a general iterative Tikhonov-Morozov method, with a general fit-to-data functional $G(p, g^\delta)$. In general, convergence rates results of *iterative Tikhonov-Morozov* methods with general fit-to-data term have not been subject of extensive research; In contrast to the non-iterative case, where we refer to [7], [17], [32] or [28] for a few references concerned with this subject.

The assertion of the above theorem, though long well known, is the key tool for the present analysis. The alternative characterization of the sequence of dual variables $\{p_n^\delta\}_{n\in\mathbb{N}}$ of the ALM as proximal point algorithm allows to apply the respective theory, and also at a later stage the analysis of flows interpolating the iterates of the dual variables of the ALM (cf. Section 7). Even more, a convergence analysis of the sequence of dual variables $\{p_n^\delta\}_{n\in\mathbb{N}}$ opens up the tools to study regularizing properties of the ALM, which is the subject of Sections 5 and 6 below.

5. Convergence Analysis

In this section we perform a convergence analysis of the ALM. The basis of this analysis is an error estimate developed by Güler in [15], which is reviewed in Proposition 5.1. Eventually, using this fundamental estimate, we are able to derive convergence rates results for the ALM. This will be done in Section 6 below. Moreover, these estimates are optimal for the particular case of the Tikhonov-Morozov regularization (6). Here, we state Güler's result:

Proposition 5.1. [15, Lem. 2.2] *Let $g^\delta \in H_2$ and set*

$$t_n := \sum_{k=1}^{n} \tau_k \,. \tag{22}$$

Then, for all $n \in \mathbb{N}$ and all $p \in H_2$

$$G(p_n^\delta, g^\delta) - G(p, g^\delta) \le \frac{\left\|p - p_0^\delta\right\|^2}{2t_n} - \frac{\left\|p - p_n^\delta\right\|^2}{2t_n} - \frac{t_n \left\|p_n^\delta - p_{n-1}^\delta\right\|^2}{2\tau_n^2}. \tag{23}$$

As a first consequence of Proposition 5.1 we derive an upper bound for the residuals in the ALM, i.e. for the term $\left\|Ku_n^\delta - g^\delta\right\|$.

Corollary 5.2. Let $g \in H_2$ be attainable and $g^\delta \in H_2$ such that $\left\|g - g^\delta\right\| \le \delta$. Then,

$$\frac{1}{2}\left\|Ku_n^\delta - g^\delta\right\|^2 \le \frac{G(p_0^\delta, g) - \inf_{p\in H_2} G(p, g)}{t_n} + \frac{\delta^2}{2}. \tag{24}$$

Proof. Setting $p = p_0^\delta$ in (23) yields

$$\frac{t_n}{2\tau_n^2}\left\|p_n^\delta - p_{n-1}^\delta\right\|^2 \le G(p_0^\delta, g^\delta) - G(p_n^\delta, g^\delta) - \frac{\left\|p_0^\delta - p_n^\delta\right\|^2}{2t_n}. \tag{25}$$

From the definition of G it follows that

$$G(p_0^\delta, g^\delta) - G(p_n^\delta, g^\delta) = G(p_0^\delta, g) - G(p_n^\delta, g) + \left\langle p_0^\delta - p_n^\delta, g - g^\delta\right\rangle \,. \tag{26}$$

Combining (26) and Young's inequality

$$\langle p, q\rangle \le \frac{\|p\|^2}{2t_n} + \frac{t_n}{2}\|q\|^2 \,, \quad p, q \in H_2$$

with (25) implies that

$$\frac{t_n}{2\tau_n^2}\left\|p_n^\delta - p_{n-1}^\delta\right\|^2 \le G(p_0^\delta, g) - G(p_n^\delta, g) + \frac{t_n}{2}\left\|g - g^\delta\right\|^2 .$$

Observing that $-G(p_n^\delta, g) \le \inf_{p\in H_2} G(p, g) < \infty$, where the second inequality follows from Lemma 4.1. This shows

$$\frac{t_n}{2\tau_n^2}\left\|p_n^\delta - p_{n-1}^\delta\right\|^2 \le G(p_0^\delta, g) - \inf_{p\in H_2} G(p, g) + \frac{t_n}{2}\left\|g - g^\delta\right\|^2 .$$

Finally, it follows from (4b) that $\left\|Ku_n^\delta - g^\delta\right\| = \tau_n^{-1}\left\|p_n^\delta - p_{n-1}^\delta\right\|$ and the assertion follows. □

Now we formulate the main theorem of this section, which states that the ALM constitutes a regularization method for the ill-posed Equation (2). For given noisy data g^δ we choose sequences $\{u_n^\delta\}_{n\in\mathbb{N}}$ and $\{p_n^\delta\}_{n\in\mathbb{N}}$ generated by the ALM (cf. Algorithm 1.1) and set

$$\mathcal{R}_n(g^\delta) := u_n^\delta \quad \text{and} \quad \mathcal{R}_n^*(g^\delta) := p_n^\delta. \tag{27}$$

Theorem 5.3. *Let $g \in H_2$ be attainable and $\{g_k\}_{k\in\mathbb{N}} \subset H_2$ be such that $\delta_k := \|g_k - g\| \to 0$ as $k \to \infty$. Further, let $\Gamma : (0,\infty) \times H_2 \to \mathbb{N}$ be such that*

$$\lim_{k\to\infty} \delta_k^2 t_{\Gamma(\delta_k,g_k)} = 0 \quad \text{and} \quad \lim_{k\to\infty} t_{\Gamma(\delta_k,g_k)} = \infty. \tag{28}$$

Then, there exists a number $c \in \mathbb{R}$ such that $\{\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k)\} \in \Lambda(c)$ for all $k \in \mathbb{N}$. In addition, each weak cluster point is a J-minimizing solution of (2) and

$$\lim_{k\to\infty} J(\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k)) = J(u^\dagger) \quad \text{and} \quad \lim_{k\to\infty} D_J^{\xi_k}(u^\dagger, \mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k)) = 0, \tag{29}$$

where $\xi_k = K^*\mathcal{R}^*_{\Gamma(\delta_k,g_k)}(g_k) \in \partial J(\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k))$. Moreover, the residuum satisfies the rate

$$\boxed{\left\|K\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k) - g\right\| = \mathcal{O}(t^{-1/2}_{\Gamma(\delta_k,g_k)}).} \tag{30}$$

Proof. Let $g^\delta \in H_2$ and set $\delta = \left\|g - g^\delta\right\|$. In the first step of the proof we derive an estimate for the sequence $\{J(u_n^\delta)\}_{n\in\mathbb{N}}$.

From (14) we know that $K^*p_n^\delta \in \partial J(u_n^\delta)$ for every $n \in \mathbb{N}$, and thus from the definition of the subgradient it follows that

$$J(u_n^\delta) \le J(u^\dagger) + \left\langle K^*p_n^\delta, u_n^\delta - u^\dagger\right\rangle = J(u^\dagger) + \left\langle p_n^\delta, Ku_n^\delta - g^\delta\right\rangle. \tag{31}$$

Using Güler's Proposition (5.1) and (26) it follows that for all $n \in \mathbb{N}$ and $p \in H_2$

$$\begin{aligned}\frac{\left\|p - p_n^\delta\right\|^2}{2t_n} &\le \frac{\left\|p - p_0^\delta\right\|^2}{2t_n} - \frac{t_n\left\|p_n^\delta - p_{n-1}^\delta\right\|^2}{2\tau_n^2} + G(p, g^\delta) - G(p_n^\delta, g_n)\\ &\le \frac{\left\|p - p_0^\delta\right\|^2}{2t_n} - \frac{t_n\left\|p_n^\delta - p_{n-1}^\delta\right\|^2}{2\tau_n^2} + G(p, g) - G(p_n^\delta, g) + \left\langle p - p_n^\delta, g - g^\delta\right\rangle.\end{aligned}$$

Then, by using Young's inequality,

$$\left\langle p - p_n^\delta, g - g^\delta\right\rangle \le \frac{1}{4t_n}\left\|p - p_n^\delta\right\|^2 + t_n\delta^2,$$

it follows that

$$\frac{\left\|p - p_n^\delta\right\|^2}{4t_n} \le \frac{\left\|p - p_0^\delta\right\|^2}{2t_n} + t_n\delta^2 - \frac{t_n}{2\tau_n^2}\left\|p_n^\delta - p_{n-1}^\delta\right\|^2 + G(p, g) - G(p_n^\delta, g).$$

Skipping the non-positive term in the previous inequality and using Lemma 4.1, which states that $-G(p_n^\delta, g) \le -\inf_{q\in H_2} G(q, g) < \infty$, shows that

$$\frac{\left\|p - p_n^\delta\right\|^2}{4t_n} \le \frac{\left\|p - p_0^\delta\right\|^2}{2t_n} + t_n\delta_n^2 + G(p, g) - \inf_{q\in H_2} G(q, g). \tag{32}$$

Now, let $\varepsilon > 0$ and choose an element $p_\varepsilon \in H_2$ such that $G(p_\varepsilon, g) \le \inf_{q\in H_2} G(q, g) + \varepsilon$. Then we conclude from (32) with the setting $p = p_\varepsilon$ that

$$\frac{\left\|p_\varepsilon - p_n^\delta\right\|}{2\sqrt{t_n}} \le \sqrt{\frac{\left\|p_\varepsilon - p_0\right\|^2}{2t_n} + t_n\delta^2 + \varepsilon}. \tag{33}$$

Set $\gamma := G(p_0, g) - \inf_{\tilde{p}\in H_2} G(\tilde{p}, g)$. Then, by using Corollary 5.2, which states that

$$\left\|Ku_n^\delta - g^\delta\right\| \le \sqrt{\frac{2\gamma}{t_n} + \delta^2}, \tag{34}$$

we obtain the desired estimate

$$J(u_n^\delta) \leq J(u^\dagger) + \left(\|p_\varepsilon\| + \sqrt{2 \left\| p_\varepsilon - p_0^\delta \right\|^2 + 4t_n^2\delta^2 + 4t_n\varepsilon} \right) \sqrt{\frac{2\gamma}{t_n} + \delta^2}. \tag{35}$$

Now, let $\Gamma : (0,\infty) \times H_2 \to \mathbb{N}$ be such that (28) is satisfied. For the sake of simplicity, we abbreviate

$$n(k) := \Gamma(\delta_k, g_k), \quad u_k := \mathcal{R}_{n(k)}(g_k) \quad \text{and} \quad p_k := \mathcal{R}^*_{n(k)}(g_k)$$

Then, by taking into account that $\delta_k \to 0$ and $t_{n(k)} \to \infty$ as $k \to \infty$ it follows from (28) and (35) that

$$\limsup_{k\to\infty} J(u_k) \leq J(u^\dagger) + 2\sqrt{2\gamma\varepsilon}.$$

Since $\varepsilon > 0$ was arbitrary, the last inequality gives that

$$\limsup_{k\to\infty} J(u_k) \leq J(u^\dagger). \tag{36}$$

Furthermore, we find from (34) and (28) that

$$\|Ku_k - g\| \leq \|Ku_k - g_k\| + \delta_k \leq \sqrt{\frac{2\gamma}{t_{n(k)}} + \delta_k^2} + \delta_k = \mathcal{O}(t_{n(k)}^{-1/2}), \tag{37}$$

which shows (30). In particular, it follows from (36) and (37) that there exists a constant $c \in \mathbb{R}$ such that

$$\sup \left(\|Ku_k\|^2 + J(u_k) \right) =: c < \infty$$

or in other words, $u_k \in \Lambda(c)$ for all $k \in \mathbb{N}$.

Consequently, according to Assumption 2.3, the sequence $\{u_k\}_{k\in\mathbb{N}}$ has a weakly convergent subsequence, say, with weak limit $\hat{u}$. Using the weak lower semi-continuity of the norm $\|\cdot\|$ and the functional J, (37) and (36) show that

$$\|K\hat{u} - g\| = 0 \quad \text{and } J(\hat{u}) \leq J(u^\dagger).$$

That is, $\hat{u}$ is a J-minimizing solution of (2) and therefore $J(\hat{u}) = J(u^\dagger)$. In particular, for each subsequence of $\{J(u_k)\}_{k\in\mathbb{N}}$ there exists a further subsequence that converges to $J(u^\dagger)$. Therefore, the first equality in (29) already holds for the whole sequence.

Finally, we find from (14) that $\xi_k = K^*p_k \in \partial J(u_k)$ and thus it follows from (35) that for all $\varepsilon > 0$

$$\begin{aligned}
0 &\leq \liminf_{k\to\infty} D_J^{\xi_k}(u^\dagger, u_k) \\
&\leq \limsup_{k\to\infty} D_J^{\xi_k}(u^\dagger, u_k) \\
&= \limsup_{k\to\infty} \left(J(u^\dagger) - J(u_k) - \left\langle K^*p_k, u^\dagger - u_k \right\rangle \right) \\
&= \limsup_{k\to\infty} \langle p_k, Ku_k - g \rangle \leq 2\sqrt{2\gamma\varepsilon}.
\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, the second equality in (29) follows. □

Theorem 5.3 shows that the ALM combined with the parameter choice (28) constitutes a *regularization method* for the ill-posed equation (2), that is, for a sequence of data g_k converging to g, the ALM approaches a J-minimizing solution.

In general, when for instance J is not strictly convex, there might exist multiple solutions of (1). We note, that the convergence result in (29) is valid for every choice of a J-minimizing solution.

Furthermore, the significance of the Bregman-distance increases with stronger convexity properties of J. For example, if J is a *total convex function*, then convergence in the Bregman-distance implies strong convergence (we refer to the work of Resmerita in [26]).

We recall that the Tikhonov-Morozov method (6) is a particular example of the ALM when the convex functional J from (5) is used. In this case Theorem 5.3 combined with Remark 2.5 implies the following corollary:

Corollary 5.4. Let $L : D(L) \subset H_1 \to H$ be a linear and closed operator with dense domain $D(L)$, and let J be as in (5). Moreover, assume that the assumptions of Theorem 5.3 are met and for $k \in \mathbb{N}$ abbreviate $u_k = \mathcal{R}_{\Gamma(\delta_k, g_k)}(g_k)$. Then,

$$\lim_{k\to\infty} \left\| Lu_k - Lu^\dagger \right\| = 0 \quad \text{and} \quad \|Ku_k - g\| = \mathcal{O}(t_{\Gamma(\delta_k,g_k)}^{-1/2})$$

for all J-minimizing solutions $u^\dagger$ of (2).

Remark 5.5. In particular Corollary 5.4 applies to the evaluation of unbounded operators. The corresponding convergence results can be found e.g. in Groetsch's book [13, Thm. 3.4]

Finally, we note that key feature of the proof of Theorem 5.3 is the fact that the descent functional $G(\cdot, g)$ is bounded from below, when g is attainable. This is exploited in order to gain an upper bound for $\{J(u_k)\}_{k\in\mathbb{N}}$ which opens the possibility to apply standard compactness arguments. However, the infimum of $G(\cdot, g)$ is not attained unless the source condition (11) is satisfied, as a consequence of which we made use of the approximate minimizers p_ε satisfying

$$G(p_\varepsilon, g) \le \inf_{p\in H_2} G(p, g) + \varepsilon, \quad \text{for all } p \in H_2$$

If the data is not attainable, the so obtained estimate (35) in the proof of Theorem 5.3 results in an arbitrarily slow speed of convergence in (29).

6. Convergence Rates

In this section we prove a convergence rate result for the iterates of the ALM under the source condition (11). The result reduces to a standard convergence rates result for Tikhonov-Morozov regularization, when J is chosen according to (5).

The following (classical) result reviews that existence of a J-minimizing solution that satisfies the source condition (11) is equivalent to the Karush-Kuhn-Tucker conditions.

Proposition 6.1 (Karush-Kuhn-Tucker)**.** *Let $g \in H_2$ be attainable, $u^\dagger \in H_1$ and $p^\dagger \in H_2$. Then the following two statements are equivalent*

(i) $u^\dagger$ is a J-minimizing solution of (2), $p^\dagger$ minimizes $G(\cdot, g)$ and

$$J(u^\dagger) + J^*(K^*p^\dagger) = \left\langle p^\dagger, g \right\rangle .$$

(ii) The Karush-Kuhn-Tucker conditions hold:

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

Proof. [8, Chap.3 Prop.4.1] □

In the following theorem we provide qualitative error estimates for the Bregman-distance of the iterates, for the residuals and the dual variables of the ALM.

Theorem 6.2. *Let $g \in H_2$ be attainable and let $g^\delta \in H_2$ satisfy (3). Assume further, that $u^\dagger$ is a J-minimizing solution of (2) that satisfies the source condition (11) with source element $p^\dagger \in H_2$. Then,*

$$\boxed{\left\| Ku_n^\delta - g \right\|^2 \le \frac{\left\| p^\dagger - p_0^\delta \right\|^2}{t_n^2} + \delta^2 \quad \textit{and} \quad D_J^{K^*p^\dagger}(u_n^\delta, u^\dagger) \le \frac{2\left\| p^\dagger - p_0^\delta \right\|^2 + 4\delta^2 t_n^2}{t_n}.} \tag{38}$$

Moreover, there exists a constant $\gamma = \gamma(p_0^\delta, p^\dagger)$ such that

$$\boxed{\left\| p_n^\delta \right\| \le \gamma + t_n\delta.} \tag{39}$$

Proof. From Proposition 6.1 it follows that $G(p^\dagger, g) \leq G(p, g)$ for all $p \in H_2$. Setting $p = p^\dagger$ in Proposition 5.1 then gives

$$\begin{aligned}\frac{\left\|p^\dagger - p_n^\delta\right\|^2}{2t_n} + \frac{t_n\left\|p_n^\delta - p_{n-1}^\delta\right\|^2}{2\tau_n^2} &\leq \frac{\left\|p^\dagger - p_0^\delta\right\|^2}{2t_n} + G(p^\dagger, g^\delta) - G(p_n^\delta, g^\delta)\\ &= \frac{\left\|p^\dagger - p_0^\delta\right\|^2}{2t_n} + G(p^\dagger, g) - G(p_n^\delta, g) + \left\langle p^\dagger - p_n^\delta, g - g^\delta\right\rangle\\ &\leq \frac{\left\|p^\dagger - p_0^\delta\right\|^2}{2t_n} + \left\langle p^\dagger - p_n^\delta, g - g^\delta\right\rangle.\end{aligned} \tag{40}$$

Below, we apply two times Young's inequality, which implies that for every $\zeta > 0$

$$\frac{\left\|p^\dagger - p_n^\delta\right\|^2}{2t_n} + \frac{t_n\left\|p_n^\delta - p_{n-1}^\delta\right\|^2}{2\tau_n^2} \leq \frac{\left\|p^\dagger - p_0^\delta\right\|^2}{2t_n} + \frac{\left\|p^\dagger - p_n^\delta\right\|^2}{2\zeta} + \frac{\zeta\delta^2}{2}. \tag{41}$$

Setting $\zeta = t_n$ in (41) and taking into account (4b) yields

$$\left\|Ku_n^\delta - g^\delta\right\|^2 = \frac{\left\|p_n^\delta - p_{n-1}^\delta\right\|^2}{\tau_n^2} \leq \frac{\left\|p^\dagger - p_0^\delta\right\|^2}{t_n^2} + \delta^2. \tag{42}$$

The choice $\zeta = 2t_n$ in (41), on the other hand, gives

$$\left\|p^\dagger - p_n^\delta\right\|^2 \leq 2\left\|p^\dagger - p_0^\delta\right\|^2 + 4t_n^2\delta^2. \tag{43}$$

From (4b) it follows that $J(u_n) - J(u^\dagger) \leq \left\langle K^* p_n^\delta, u_n - u^\dagger\right\rangle$ and consequently, it follows that

$$D_J^{K^*p^\dagger}(u_n, u^\dagger) = J(u_n) - J(u^\dagger) - \left\langle K^*p^\dagger, u_n - u^\dagger\right\rangle \leq \left\langle p_n^\delta - p^\dagger, Ku_n - g\right\rangle.$$

Applying again Young's inequality and combining the estimates (42) and (43) we find that for $\eta > 0$ that

$$\begin{aligned}D_J^{K^*p^\dagger}(u_n, u^\dagger) &\leq \frac{\left\|p_n^\delta - p^\dagger\right\|^2}{2\eta} + \frac{\eta\left\|Ku_n - g\right\|^2}{2}\\ &\leq \frac{\left\|p_n^\delta - p^\dagger\right\|^2}{2\eta} + \eta\left\|Ku_n - g^\delta\right\|^2 + \eta\delta^2\\ &\leq \left\|p^\dagger - p_0^\delta\right\|^2\left(\frac{1}{\eta} + \frac{\eta}{t_n^2}\right) + 2\delta^2\left(\frac{t_n^2}{\eta} + \eta\right).\end{aligned}$$

The right hand side is minimized for $\eta = t_n$, which finally shows the assertion. $\square$

Theorem 6.2 can be used to prove convergence rates for the ALM.

Theorem 6.3. *Assume that $g \in H_2$ is attainable and assume that $\{g_k\}_{k\in\mathbb{N}}$ is a sequence in H_2 such that $\|g_k - g\| =: \delta_k \to 0$ as $k \to \infty$. Moreover, assume that $\Gamma : (0,\infty) \times H_2 \to \mathbb{N}$ is such that*

$$\lim_{k\to\infty} t_{\Gamma(\delta_k, g_k)} = \infty.$$

Then, the following two conditions are equivalent:

(i) There exists a J-minimizing solution $u^\dagger \in H_1$ of (2) that satisfies the source condition (11) with a source element $p^\dagger \in H_2$ and there exists $C \in \mathbb{R}$ such that

$$\delta_k t_{\Gamma(\delta_k, g_k)} \leq C. \tag{44}$$

(ii) For $k \to \infty$ we have

$$\left\|K\mathcal{R}_{\Gamma(\delta_k, g_k)}(g_k) - g\right\| = \mathcal{O}(t^{-1}_{\Gamma(\delta_k, g_k)}) \quad \text{and} \quad \left\|\mathcal{R}^*_{\Gamma(\delta_k, g_k)}(g_k)\right\| = \mathcal{O}(1). \tag{45}$$

Additionally, if (i) or (ii) holds, it follows that

$$D_J^{K^*p^\dagger}(\mathcal{R}_{\Gamma(\delta_k, g_k)}(g_k), u^\dagger) = \mathcal{O}(t^{-1}_{\Gamma(\delta_k, g_k)}) \tag{46}$$

and each weak cluster point of $\left\{\mathcal{R}^*_{\Gamma(\delta_k, g_k)}(g_k)\right\}_{k\in\mathbb{N}}$ is a minimizer of $G(\cdot, g)$.

Proof. Througout the proof we assume that $\{u_n^\delta\}_{n\in\mathbb{N}}$ and $\{p_n^\delta\}_{n\in\mathbb{N}}$ are two sequences generated by the ALM w.r.t. generic data $g^\delta \in H_2$ and we set $\delta = \|g^\delta - g\|$. Moreover, we use the abbreviations

$$u_k = \mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k) \quad \text{and} \quad p_k = \mathcal{R}^*_{\Gamma(\delta_k,g_k)}(g_k).$$

Assume that item (i) holds. Then (45) and (46) follow from Theorem 6.2. In particular, the sequence $\{p_k\}_{k\in\mathbb{N}}$ is bounded. We prove that each of its weak cluster points is a minimizer of $G(\cdot;g)$. To this end, obserfve that the optimality condition for (21) gives

$$\tau_n^{-1}(p_{n-1}^\delta - p_n^\delta) \in \partial G(\cdot,g^\delta)(p_n) = \partial G(\cdot,g)(p_n^\delta) - (g^\delta - g).$$

Together with the update rule (4b) in ALM this gives

$$Ku_n^\delta - g \in \partial G(\cdot,g)(p_n^\delta). \tag{47}$$

Now choose a weakly convergent sub-sequence $\{p_{\rho(k)}\}_{k\in\mathbb{N}}$ with limit $\hat{p}$. Then we find from the weak-strong closedness of $\partial G(\cdot,g)$ (cf. [8, Chap I. Cor. 5.1]) and (47) (we replace u_n^δ by $u_{\rho(k)}$ and p_n^δ by $p_{\rho(k)}$) that

$$0 = \lim_{k\to\infty} Ku_{\rho(k)} - g \in \partial G(\cdot,g)(\hat{p}).$$

This shows that $\hat{p}$ is a minimizer of $G(\cdot,g)$.

Now let (ii) hold. We first note, that setting $p = p_n^\delta$ in Proposition 5.1 yields

$$\left\|Ku_n^\delta - g^\delta\right\| = \tau_n^{-1}\left\|p_n^\delta - p_{n-1}^\delta\right\| \le t_n^{-1}\left\|p_n^\delta - p_0^\delta\right\|. \tag{48}$$

and therefore

$$t_n\delta = t_n\left\|g - g^\delta\right\| \le t_n(\left\|Ku_n^\delta - g^\delta\right\| + \left\|Ku_n^\delta - g\right\|) = \left\|p_n^\delta - p_0^\delta\right\| + t_n\left\|Ku_n^\delta - g\right\|. \tag{49}$$

Replacing g^δ by g_k, δ by δ_k and n by $\Gamma(\delta_k,g_k)$ as well as u_n^δ by u_k and p_n^δ by p_k in (49) shows together with (45) that

$$\delta_k t_{\Gamma(\delta_k,g_k)} \le \left\|p_k - p_0^\delta\right\| + t_{\Gamma(\delta_k,g_k)}\|Ku_k - g\| = \mathcal{O}(1).$$

That is, there exsist $C \in \mathbb{R}$ such that (44) holds.

Next, we find from (14) that $K^*p_k \in \partial J(u_k)$ and thus for all $u \in H_1$

$$J(u_k) \le J(u) + \langle K^*p_k, u - u_k\rangle \le J(u) + \|p_k\|\,\|Ku_k - Ku\|. \tag{50}$$

The estimates (48) and (50) imply that $u_k \in \Lambda(c)$ for an appropriate constant $c \in \mathbb{R}$ and all $k \in \mathbb{N}$. Thus (with the same argumentation as in the proof of Theorem 5.3) we can select a sub-sequence that weakly converges to a J-minimizing solution $u^\dagger$ of (2). Denoting the subsequence again by $\{u_k\}_{k\in\mathbb{N}}$ one finds

$$\lim_{k\to\infty} J(u_k) = J(u^\dagger). \tag{51}$$

Finally, it follows from (48) that $Ku_k - g \to 0$ strongly in H_2 as $k \to \infty$. Relation (47) and the arguments thereafter show that each weak cluster point $p^\dagger$ of $\{p_k\}_{k\in\mathbb{N}}$ is a minimizer of $G(\cdot,g)$. This implies that

$$G(p_k,g_k) - G(p^\dagger,g) \ge \langle p_k, g - g_k\rangle.$$

Moreover, by setting $p = p^\dagger$ in Proposition 5.1 and neglecting non-positive terms in the right hand side of Güler's inequality we find

$$G(\delta_k,g_k) - G(p^\dagger,g) \le \frac{\left\|p^\dagger - p_k\right\|^2}{2t_k} + \left\langle p^\dagger, g - g_k\right\rangle.$$

The previous two inequalities show that

$$\lim_{k\to\infty} J^*(K^*(p_k)) - \langle p_k, g_k\rangle = \lim_{k\to\infty} G(p_k,g_k) = G(p^\dagger,g) = J^*(K^*p^\dagger) - \left\langle p^\dagger, g\right\rangle. \tag{52}$$

Since $K^*p_k \in \partial J(u_k)$ it follows that $J(u_k) + J^*(K^*p_k) = \langle u_k, K^*p_k\rangle$ and thus

$$J(u_k) + J^*(K^*p_k) - \langle p_k, g_k\rangle = \langle Ku_k - g_k, p_k\rangle.$$

Passing to the limit $k \to \infty$, this equality together with (51) and (52) gives

$$J(u^\dagger) + J^*(K^*p^\dagger) = \left\langle p^\dagger, g\right\rangle = \left\langle K^*p^\dagger, u^\dagger\right\rangle.$$

This, however, is equivalent to $K^*p^\dagger \in \partial J(u^\dagger)$, that is, $u^\dagger$ satisfies the source condition with source element $p^\dagger$. $\square$

Theorem 6.3 states that for each parameter choice rule Γ satisfying (44), the residuals converge with a rate of t_k^{-1} and the sequence of dual iterates in the ALM is bounded, or in other words, converge *slower* by a rate t_k^{-1} than the residuals. In fact, it turns out that these two assertions are equivalent.

We close this section by applying the result in Theorem 6.2 to the Tikhonov-Morozov method (6), that is, we choose the regularization functional J as in (5). Then Theorem 6.2 reads (cf. Remark 2.5):

Corollary 6.4. Let $g \in H_2$ be attainable and let $g^\delta \in H_2$ be such that $\left\|g - g^\delta\right\| = \delta$. Assume further that $u^\dagger$ is a J-minimizing solution of (2) that satisfies the source condition (11) with source element $p^\dagger \in H_2$. Then,

$$\left\|Ku_n^\delta - g^\delta\right\|^2 \leq \frac{\left\|p^\dagger - p_0^\delta\right\|^2}{t_n^2} + \delta^2 \quad \text{and} \quad \left\|Lu_n^\delta - Lu^\dagger\right\|^2 \leq \frac{2\left\|p^\dagger - p_0^\delta\right\|^2 + 4\delta^2 t_n^2}{t_n}. \tag{53}$$

Moreover, there exists a constant $\gamma = \gamma(p_0^\delta, p^\dagger)$ such that

$$\left\|L^*Lu_n^\delta\right\| \leq \gamma + \|K^*\| t_n \delta. \tag{54}$$

It is well known that the classical iterated Tikhonov-Morozov regularization can converge with order $\delta^{1-\epsilon}$, $\epsilon > 0$ under appropriate assumptions on the solution (see [13]). Such results are not available for the general ALM, and consequently Chuck Groetsch's benchmarks have not been reached so far.

The convergence rates in Theorem 6.2 were already proven (under the same assumptions) by Burger et al. in [6], however for the stationary case $\tau_n \equiv \tau$. Our results, in addition provide an equivalence relation between the standard assumptions (parameter choice rule and source condition) and the boundedness of the sequence of dual iterates.

In our opinion, the speed of convergence of the dual sequence in the ALM could be the key in order to reach the benchmark results of Chuck Groetsch in the special case of Tikhonov-Morozov regularization. Conditions on the solutions of (2) that guarantee faster convergence of the dual sequence of the ALM will therefore be subject of further studies.

7. Evolution Equations

In this section we study the following system of evolution equations

$$\begin{aligned} p'(t) &= g^\delta - Ku(t), && (55a)\\ K^*p(t) &\in \partial J(u(t)), && (55b)\\ p(0) &= p_0^\delta. && (55c)\end{aligned}$$

and its relation to the ALM (cf. Definition 1.1). With $K = \mathrm{Id}$, equation (55), has been proposed in [5] and finds applications for image denoising.

For the special case $K = \mathrm{Id}$ existence of a solution of (55) has been proven in [4] (see also [12]). In general, for bounded linear operators K, existence of a solution of (55), as well as the relation to the ALM has been studied in the first author's thesis [11]. Here, we present a summary of the most important results.

In the following we study the connection of Equation (55) and the ALM, which for the sake of simplicity is assumed with constant stepsize $\tau > 0$. The corresponding results for the non-stationary case were proven in [11].

Let $g, p_0^\delta \in H_2$ and denote by (u_n^δ, p_n^δ) the n-th iterate of the ALM. For $t \in [(n-1)\tau, n\tau)$ we define

$$u_\tau(t) = u_n^\delta \tag{56}$$

$$p_\tau(t) = \tau^{-1}\left(t - (n-1)\tau)p_n^\delta + (n\tau - t)p_{n-1}^\delta\right). \tag{57}$$

In other words, $u_\tau(t)$ and $p_\tau(t)$ are the piecewise constant and piecewise affine interpolations of the sequences $\{u_n^\delta\}_{n\in\mathbb{N}}$ and $\{p_n^\delta\}_{n\in\mathbb{N}}$ of the ALM.

The function $p_\tau(t)$ is differentiable almost everywhere and satisfies $p_\tau(0) = p_0^\delta$ and we find by (4b) that for all $t \in ((n-1)\tau, n\tau)$

$$p_\tau'(t) = \tau^{-1}(p_n^\delta - p_{n-1}^\delta) = g^\delta - Ku_n = g^\delta - Ku_\tau(t). \tag{58}$$

Moreover, we find from Theorem 3.1 that

$$K^* p_\tau(n\tau) = K^* p_n^\delta \in \partial J(u_n) = \partial J(u_\tau(t)). \tag{59}$$

The considerations in (58) and (59) show, that u_τ and p_τ *almost* satisfy (55). We show that u_τ and p_τ converge to solutions of (55) as $\tau \to 0$.

Due to the characterization of the sequence $\{p_n^\delta\}_{n\in\mathbb{N}}$ by the proximal point algorithm (cf. Proposition 4.2), we are able to apply classical results of semi-group theory in order to show convergence of the piecewise affine functions p_τ to a *strong solution* of

$$p'(t) = -\partial^0 G(p(t), g^\delta) \tag{60a}$$
$$p(0) = p_0^\delta. \tag{60b}$$

Here, $\partial^0 G(p,g)$ denotes the unique element of $\partial G(p,g)$ with minimal norm (presumably $\partial G(p,g) \neq \emptyset$). A strong solution of (60) is a absolutely continuous function $p : [0,\infty) \to H_2$ such that (60a) is satisfied almost everywhere and that $p(t) \to p_0^\delta$ as $t \to 0^+$.

Proposition 7.1. *Let $g^\delta, p_0^\delta \in H_2$ and assume that p_0^δ satisfies* (17)*, i.e. there exists $u_0^\delta \in H_1$ such that $K^* p_0^\delta \in \partial J(u_0^\delta)$. Then*

(i) There exists a unique strong solution $p : [0,\infty) \to H_2$ of (60)*.*
(ii) The piecewise affine interpolations p_τ converge uniformly to p and

$$\|p_\tau(t) - p(t)\| \leq \frac{\tau}{\sqrt{2}} \left\|\partial^0 G(p_0^\delta, g^\delta)\right\|. \tag{61}$$

(iii) The function p is Lipschitz-continuous with Lipschitz-constant $c_L = \left\|Ku_0^\delta - g^\delta\right\|$.

Proof. Since $K^* p_0^\delta \in \partial J(u_0^\delta)$ it follows that $Ku_0^\delta \in \partial(J^* \circ K^*)(p_0^\delta)$ and thus according to the definition of G in (19)

$$Ku_0^\delta - g^\delta \in \partial(J^* \circ K^*)(p_0^\delta) - g^\delta = \partial G(p_0^\delta, g^\delta).$$

In other words, $p_0^\delta \in D(\partial G(\cdot, g^\delta))$ and hence (i) follows from [3, Thm. 3.1]. Item (ii) follows from [23, Thm. 3.20].

It remains to prove the third item. As noted above, the functions $p_\tau(t)$ are differentiable for almost all $t \geq 0$ and p_τ'satisfies (58). From Corollary 3.3 it follows that $p_\tau'(t)$ is non-increasing and since p_0^δ satisfies (17) we find that (cf. Corollary 3.3)

$$\|p_\tau'(t)\| \leq \left\|Ku_0^\delta - g^\delta\right\| = c_L.$$

Thus, $p_\tau(t)$ is Lipschitz-continuous with constant c_L and it follows from (2) that for $s,t \geq 0$ and all $\tau > 0$

$$\|p(s) - p(t)\| \leq \|p_\tau(s) - p(s)\| + \|p_\tau(t) - p(t)\| + \|p_\tau(s) - p_\tau(t)\| \leq \sqrt{2}\tau \left\|\partial^0 G(p_0^\delta, g^\delta)\right\| + c_L \|s-t\|.$$

Since the equation holds for all $\tau > 0$, the assertion follows by taking $\tau \to 0+$. □

Assertion (2) in Proposition 7.1 states that the piecewise affine interpolations $p_\tau(t)$ of the sequence $\{p_n^\delta\}_{n\in\mathbb{N}}$ converge uniformly on $[0,\infty)$, that is

$$\lim_{\tau\to 0^+} \sup_{t\geq 0} \|p_\tau(t) - p(t)\| = 0.$$

In the thesis [11, Cor. 3.3.6] we proved the slightly stronger result:

Proposition 7.2. *Let $g^\delta, p_0^\delta \in H_2$ and assume that p_0^δ satisfies* (17). *Moreover, let p be the unique solution of* (7.1). *Then, for all $T > 0$*

$$\lim_{\tau\to 0^+} \int_0^T \|p_\tau(t) - p(t)\|^2 + \|p'_\tau(t) - p'(t)\|^2 \,\mathrm{d}t = 0.$$

We note that Proposition 7.2 states that the sequence of functions $\{t \mapsto \|p_\tau(t) - p(t)\|\}_{\tau\geq 0}$ converges to zero in $\mathrm{W}^{1,2}(0,T)$. This already implies uniform convergence by the continuous embedding $\mathrm{W}^{1,2}(0,T) \hookrightarrow C(0,T)$ ([1, Thm. 5.4] with $n = 1$, $m = 1$ and $p = 2$).

By using Proposition 7.2 we can prove the main theorem of this section, which provides estimates for the errors of the interpolates to the true solution.

Theorem 7.3. *Let $g^\delta \in H_2$ and assume that $p_0^\delta \in H_2$ satisfies* (17). *Moreover, let p be the unique solution of* (7.1). *Then*

(i) There exists $u : [0,\infty) \to H_1$ such that u and p satisfy (55) *for almost all $t \geq 0$.*
(ii) For all $t \geq 0$ we have

$$D_J^{K^*p(t)}(u_\tau(t), u(t)) \leq \tau(2 + \sqrt{2}) \left\|Ku_0^\delta - g^\delta\right\|^2 . \tag{62}$$

Proof. Let $t > 0$ and choose $n \in \mathbb{N}$ such that $t \in ((n-1)\tau, n\tau]$. With this choice it follows from Corollary 3.3 and the subsequent remark, that

$$\left\|Ku_\tau(t) - g^\delta\right\| = \left\|Ku_n - g^\delta\right\| \leq \left\|Ku_0^\delta - g^\delta\right\|$$

This shows that $\|Ku_\tau(t)\|$ is uniformly bounded, let us say $\|Ku_\tau(t)\| \leq c_1$ for all $\tau > 0$. Moreover, we find from (59) and the definition of a subgradient that for all $u \in H_1$

$$\begin{aligned} J(u_\tau(t)) &\leq J(u) + \langle K^*p_\tau(n\tau), u_\tau(t) - u\rangle \\ &= \langle p_\tau(n\tau) - p(n\tau), Ku_\tau(t) - Ku\rangle + \langle p(n\tau), Ku_\tau(t) - Ku\rangle . \end{aligned}$$

We choose $u \in D(J)$. Then it follows from (61) that

$$J(u_\tau(t)) \leq J(u) + (\|K(u_\tau(t))\| + \|Ku\|)\left(\|p(\tau n)\| + \frac{\tau}{\sqrt{2}}\left\|\partial^0 G(p_0^\delta, g^\delta)\right\|\right).$$

Since $p(t)$ is Lipschitz-continuous with constant c_L as in Proposition 7.1 we find that

$$\|p(t) - p(n\tau)\| \leq c_L(n\tau - t) \leq c_L\tau \tag{63}$$

and consequently, for τ sufficiently small, there exists a constant $c_2 \in \mathbb{R}$ such that

$$J(u_\tau(t)) \leq J(u) + (c_1 + \|Ku\|)\left(c_L\tau + \|p(t)\| + \frac{\tau}{\sqrt{2}}\left\|\partial^0 G(p_0^\delta, g^\delta)\right\|\right) \leq c_2 < \infty .$$

In other words, let $t > 0$ arbitrary but fixed, then $u_\tau(t) \in \Lambda(c_1 + c_2)$ and therefore from Assumption 2.3 it follows that there exists a sequence $\tau_k \to 0^+$ and an element $u(t) \in H_1$ such that $u_{\tau_k}(t) \rightharpoonup u(t)$.

Now, we show that u and p solve (55). First we note that from (61) and (63) it follows that $p_\tau(n\tau) \to p(t)$ as $\tau \to 0^+$ (and $n \to \infty$ accordingly). Setting $\tau = \tau_k$ it follows from the strong-weak closedness of ∂J (cf. [8, Chap.I Cor. 5.1]) and (59) that

$$K^*p(t) = \lim_{k\to\infty} K^*p_{\tau_k}(n\tau_k) \in \partial J(\underset{k\to\infty}{w\text{-}\lim}\, u_{\tau_k}(t)) = \partial J(u(t)).$$

Secondly, we note that Proposition (7.2) implies that for all $T > 0$ the sequence p'_τ of derivatives converges to p' with respect to the strong topology of $\mathrm{L}^2(0,T,H_2)$, that is

$$\lim_{k\to\infty} \int_0^T \left\|p'_{\tau_k}(t) - p'(t)\right\|^2 \mathrm{d}t = 0.$$

It follows from [9, Chap. 1.3. Thm. 5] that we can select a subsequence indexed by $\rho(k)$ such that $p'_{\tau_{\rho(k)}}(t) \to p'(t)$ strongly for almost all $t \in [0,T]$. It finally follows from (58) that

$$p'(t) = \lim_{k\to\infty} p'_{\tau_{\rho(k)}}(t) = w\text{-}\lim(g - Ku_{\tau_{\rho(k)}}(t)) = g^\delta - Ku(t). \tag{64}$$

In order to prove (ii) we assume that $u(t)$ and $p(t)$ solve (55). Since $\|p_\tau'(t)\| = \left\|Ku_\tau(t) - g^\delta\right\|$ is non-increasing and uniformly bounded by $\left\|Ku_0^\delta - g^\delta\right\|$, it follows from (64) that the same holds for $\|p'(t)\| = \left\|Ku(t) - g^\delta\right\|$. Moreover, it follows from (59) that $K^*p_\tau(n\tau) \in \partial J(u_\tau(t))$. The later and the definition of the subgradient imply that

$$J(u_\tau(t)) - J(u(t)) \leq \langle K^*p_\tau(n\tau), u_\tau(t) - u(t)\rangle\,.$$

In summary, we find that

$$\begin{aligned}
D_J^{K^*p(t)}(u_\tau(t), u(t)) &= J(u_\tau(t)) - J(u(t)) - \langle K^*p(t), u_\tau(t) - u(t)\rangle \\
&\leq \langle p_\tau(n\tau) - p(t), K(u_\tau(t) - u(t))\rangle \\
&\leq \|p_\tau(n\tau) - p(t)\| \left(\left\|Ku_\tau(t) - g^\delta\right\| + \left\|Ku(t) - g^\delta)\right\|\right) \\
&\leq 2\,\|p_\tau(n\tau) - p(t)\| \left\|Ku_0^\delta - g^\delta\right\|\,.
\end{aligned}$$

Moreover, it follows from (61) and the Lipschitz-continuity of p_τ (with constant $c_L = \left\|Ku_0^\delta - g\right\|$) that

$$\|p_\tau(n\tau) - p(t)\| \leq \|p_\tau(n\tau) - p_\tau(t)\| + \|p_\tau(t) - p(t)\| \leq \tau(c_L + 1/\sqrt{2}\left\|\partial^0 G(p_0^\delta, g)\right\|).$$

As shown in the proof of Proposition 7.1, $Ku_0^\delta - g \in \partial G(p_0^\delta, g)$ and thus $\left\|\partial^0 G(p_0^\delta, g)\right\| \leq \left\|Ku_0^\delta - g\right\|$ and the assertion follows. □

In contrast to (60), Equation (55) in general admits multiple solutions. However, solutions are unique modulo $\ker(K)$ and are uniformly continuous with respect to the Bregman distance. These are the assertions of the following two theorems:

Theorem 7.4. *Let p_0^δ satisfy* (17)*, i.e. there exists u_0^δ such that $K^*p_0^\delta \in \partial J(u_0^\delta)$ and assume that p and u satisfy* (55)*. Then, for all $0 \leq s, t$*

$$D_J^{K^*p(t)}(u(s), u(t)) \leq 2\left\|Ku_0^\delta - g^\delta\right\|^2 |s - t|\,.$$

Proof. According to (55b) we find $K^*p(s) \in \partial J(u(s))$ and therefore it follows that $J(u(s)) - J(u(t)) \leq \langle K^*p(s), u(s) - u(t)\rangle$. Thus, Proposition 7.1 (3) and the monotonicity of $\left\|Ku(t) - g^\delta\right\|$ give

$$\begin{aligned}
D_J^{K^*p(t)}(u(s), u(t)) &\leq J(u(s)) - J(u(t)) - \langle K^*p(t), u(s) - u(t)\rangle \\
&= \left\langle p(s) - p(t), Ku(s) - g^\delta\right\rangle - \left\langle p(s) - p(t), Ku(t) - g^\delta\right\rangle \\
&\leq \|p(s) - p(t)\| \left(\left\|Ku(s) - g^\delta\right\| + \left\|Ku(t) - g^\delta\right\|\right) \\
&\leq 2\left\|Ku_0^\delta - g^\delta\right\|^2 |s - t|\,.
\end{aligned}$$

□

Theorem 7.5. *Let p be the unique solution of* (7.1) *and assume that $u_1, u_2 : [0, \infty) \to \infty$ are such that $p(t)$ and $u_i(t)$ satisfy* (55) *for $i = 1, 2$. Then, for all $t \geq 0$*

$$Ku_1(t) = Ku_2(t) \quad \textit{and} \quad J(u_1(t)) = J(u_2(t)). \tag{65}$$

Proof. The equality $Ku_1(t) = Ku_2(t)$ follows directly from (55a). Now, let $t \geq 0$. Then we find that

$$\begin{aligned}
D_J^{K^*p(t)}(u_1(t), u_2(t)) &= J(u_1(t)) - J(u_2(t)) - \langle K^*p(t), u_1(t) - u_2(t)\rangle \\
D_J^{K^*p(t)}(u_2(t), u_2(t)) &= J(u_2(t)) - J(u_1(t)) - \langle K^*p(t), u_2(t) - u_1(t)\rangle\,.
\end{aligned}$$

Adding these equalities gives

$$D_J^{K^*p(t)}(u_1(t), u_2(t)) + D_J^{K^*p(t)}(u_2(t), u_2(t)) = 0.$$

Since the Bregman-distance is non-negative, it follows that

$$0 = D_J^{K^*p(t)}(u_1(t), u_2(t)) = J(u_1(t)) - J(u_2(t)) - \langle p(t), Ku_1(t) - Ku_2(t)\rangle\,.$$

As noted above, we have $Ku_1(t) = Ku_2(t)$ and therefore $J(u_1(t)) = J(u_2(t))$. □

Remark 7.6. (i) If the restriction of J on each of the sets $u+\ker(K)$ is strictly convex, then it follows from Theorem 7.5, that (55) admits a unique solution (u,p). In this case, it follows from the construction of u in the proof of Proposition 7.3 that $u_\tau(t) \rightharpoonup u(t)$ for $\tau \to 0^+$. Ja

(ii) It J is strictly convex on H_1, it follows that ∂J^* contains at most one element. This can be seen as follows: If $p_0 \in H_2$ and $u_0 \in \partial J^*(p_0)$, then

$$J^*(p_0) = \langle u_0, p_0\rangle - J(u_0) = \sup_{u\in H_1} \langle u, p_0\rangle - J(u)$$

Since J is strictly convex, $u \mapsto \langle u, p_0\rangle - J(u)$ is strictly concave and hence u_0 is the unique maximum. With a slight abuse of notation, let us denote $\partial J^*(p_0) = \nabla J^*(p_0)$ (note, however, that J^* does not need to be differentiable).

Consequently it follows from (10) that $K^*p \in \partial J(u) \Leftrightarrow u = \nabla J^*(K^*p)$, and Equation (55) can be rewritten as

$$p'(t) = g^\delta - \nabla(J^* \circ K^*)(p(t)) \quad \text{and} \quad p(0) = p_0^\delta.$$

and $u(t) = \nabla J^*(K^*p(t))$. In other words, $u(t)$ is uniquely determined by $p(t)$.

(iii) The case where J is a convex regulariation functional with *linear growth* and $K = \mathrm{Id}$ is a typical situation in image denoising problems and a particular instance where (i) is satisfied but not (ii). The special example, where J is chosen to be the *total variation semi-norm* on $H_1 = \mathrm{L}^2(\Omega)$ was studied e.g. in [4] and [11].

Remark 7.7. For a solutions u of (55) the element $u(t)$ at a finite time $t \geq 0$ can be considered as a regularized solution of (2). For the special case where $J(u) = \frac{1}{2}\|u\|^2$, Equation (55) can be written as

$$u' + K^*Ku = K^*g^\delta \quad \text{and} \quad u(0) = u_0^\delta$$

which is known as *Showalter's method* (cf. [30, 31]). Equation (55) therefore can be considered as a generalization of this method.

For the sake of lucidity, we restrict our analysis on the existence of solutions of (55) and their basic properties (uniqueness, continuity etc.) and do not discuss their regularizing properties. We note, that the convergence (rates) results in Sections 5 and 6 can be transferred to the time-continuous case where the time parameter t playes the role of t_n. The results were proven by Burger et al. in [6, Thm. 5.2].

7.1. **The Tikhonov-Morozov Method.** In the special setting of Tikhonov-Morozov regularization, i.e. when J is as in (5), we have $\partial J(u) = L^*Lu$. Therefore, Equation (55) reads as

$$\boxed{(L^*Lu(t))' = g^\delta - Ku(t) \quad \text{and} \quad L^*Lu(0) = p_0^\delta} \tag{66}$$

for an initial element $p_0^\delta \in H_2$. For the special case of $K = \mathrm{Id}$, Equation (66) has been studied by Groetsch & Scherzer in [14]. Applying Theorems 7.3 to 7.5 gives the following stability estimates

Corollary 7.8. Let $p_0^\delta \in \mathrm{ran}(L^*L)$. Then (66) has at least on solution $u : [0,\infty) \to H_1$ and for all $s,t \geq 0$ we have

$$\|L(u_\tau(t) - u(t))\|^2 \leq \tau(2+\sqrt{2})\left\|Ku_0 - g^\delta\right\|^2$$

and

$$\|L(u(s) - u(t))\|^2 \leq 2\|Ku_0 - g\|^2 |s-t|$$

If additionally $\ker(L) \cap \ker(K) = \{0\}$, then u is unique.

We close this section by an example that was studied by Groetsch & Scherzer in [14]. There, the authors considered the problem of the stable evaluation of the derivative of a smooth function given a noisy (non-differentiable) approximation.

Example 7.9. Let $\Omega \subset \mathbb{R}^N$ be a open and bounded domain with smooth boundary $\partial\Omega$ and $H_1 = H_2 = \mathrm{L}^2(\Omega)$ as well as $H = \mathrm{L}^2(\Omega, \mathbb{R}^N)$ and set $L = \nabla$ with $\mathrm{H}^1(\Omega) = D(L) \subset \mathrm{L}^2(\Omega)$. Then L is linear, closed and densely defined (with respect to the L^2-topology). Moreover we find from [2, pp.63] that

$$D(L^*L) = D(\partial J) = \left\{u \in \mathrm{H}^2(\Omega) \ : \ \nabla u \cdot \nu = 0,\ \mathcal{H}^{N-1}\text{-a.e. on } \partial\Omega\right\}$$

and $\partial J(u) = -\Delta u$ (ν denotes the outer unit normal vector on $\partial\Omega$). further, we set $K = \mathrm{Id}$ and $p_0^\delta = 0 \in \mathrm{ran}(L^*L)$. Then, the evolution equation (66) turns out to be the third order equation

$$\boxed{\frac{d}{dt}\Delta u(t,x) = u(t,x) - g^\delta(x) \quad \text{for all } x \in \Omega,} \tag{67}$$

with homogenuous Neumann boundary condition

$$\nabla u(x) \cdot \nu(x) = 0 \quad \text{for } \mathcal{H}^{N-1}\text{-a.e. } s \in \partial\Omega, \tag{68}$$

and initial conditions

$$\Delta u(0,x) = 0. \tag{69}$$

We show that the unique solution u of the third order equation (67) is Hölder-continuous (in time) with exponent $\alpha = 1/2$ and that the piecewise constant interpolations u_τ of $\left\{u_n^\delta\right\}_{n\in\mathbb{N}}$ approximate u with order $\tau^{1/2}$.

To this end, we note that by Green's formula

$$\int_\Omega u(t,x) - g^\delta(x)\,\mathrm{d}x = \int_\Omega \frac{d}{dt}\Delta u(t,x)\,\mathrm{d}x = \frac{d}{dt}\int_{\partial\Omega} \nabla u(t,x)\cdot\nu\,\mathrm{d}x = 0. \tag{70}$$

Hence it follows that

$$\int_\Omega u(s,x) - u(t,x)\,\mathrm{d}x = 0 \text{ for all } s,t \geq 0\,.$$

Consequently there exists a embedding constant $C = C(\Omega)$ such that (cf. [35, Thm.4.2.1])

$$\|u(s) - u(t)\|_{\mathrm{L}^2} \leq C\,\|\nabla(u(s) - u(t))\|_{\mathrm{L}^2(\Omega,\mathbb{R}^N)}$$

This together with Corollary 7.8 then shows that

$$\|u(s) - u(t)\|_{\mathrm{L}^2} \leq \sqrt{2}C\,\left\|g^\delta\right\|_{\mathrm{L}^2}^2 \sqrt{|s-t|}.$$

In particular, the continuity of u and (70) imply

$$\int_\Omega u(0,x)\,\mathrm{d}x = \int_\Omega \lim_{t\to 0^+} u(t,x)\,\mathrm{d}x = \lim_{t\to 0^+}\int_\Omega u(t,x)\,\mathrm{d}x = \int_\Omega g^\delta(x)\,\mathrm{d}x.$$

Since $\Delta u(0,x) = 0$ and $\nabla u(0,x)\cdot\nu(x) = 0$ on $\partial\Omega$, $u(0,x)$ is constant and the constant is given by

$$u(0,x) = \frac{1}{\lambda^N(\Omega)}\int_\Omega g^\delta(x)\,\mathrm{d}x =: \bar{g}^\delta.$$

Next, we apply the error estimate in Theorem 7.3 (2) to the present example. Here, we find from (4b) that

$$g^\delta - u_n = \tau^{-1}\Delta(u_{n-1}^\delta - u_n).$$

Again, by Green's formula one finds as in (70)

$$\bar{g}^\delta = \int_\Omega u_n(x)\,\mathrm{d}x\,.$$

Consequently we find

$$\int_\Omega u_\tau(t,x) - u(t,x)\,\mathrm{d}x = \int_\Omega u_\tau(t,x) - g^\delta(x)\,\mathrm{d}x + \int_\Omega g^\delta(x) - u(t,x)\,\mathrm{d}x = 0$$

for all $t \geq 0$. Therefore, using the embedding constant C from above and Corollary 7.8 it follows that

$$\|u_\tau(t) - u(t)\|_{\mathrm{L}^2}^2 \leq C\,\|\nabla(u_\tau(t) - u(t))\|_{\mathrm{L}^2(\Omega,\mathbb{R}^N)}^2 \leq \tau C(2+\sqrt{2})\int_\Omega \left|g^\delta(x) - \bar{g}^\delta\right|^2\,\mathrm{d}x.$$

8. Conclusion

This paper is concerned with the application of the *Augmented Lagrangian method (ALM)* to the solution of linear inverse problems. This means, we interpret the iterates of the ALM as regularized solutions of the ill-posed linear equation

$$Ku = g$$

where a noisy righthand side g^δ serves as an input data. We perform convergence studies for ALM with general regularization functionals J, and prove convergence of the iterates to a *J-minimizing solution* $u^\dagger$ of the linear equation as the noise level decreases.

The motivation for this approach is due to the fact that the *Tikhonov-Morozov method*, which gained much attraction in the field of stable evaluation of unbounded linear operators L, is a special instance of the ALM for the choice

$$J(u) = \frac{1}{2}\|Lu\|^2 .$$

Thus, benchmark results concerning convergence (rates) results are available, many of which were shown by Chuck Groetsch. We refer to the excellent monograph [13] for a rich collection of these results.

Additionally, we remark that the ALM is equivalent to *Bregman distance regularization*, which attained much attention in the image processing community recently.

Unlike most of the approaches so far, we base our analysis on duality arguments from convex analysis. In doing so, we are able to reprove convergence and convergence rates results, already known for *Bregman distance regularization*, and derive convergence rates for the dual variables. In particular, we can prove that under the standard source-condition

$$K^*p^\dagger \in \partial J(u^\dagger), \qquad \text{for a source-element } p^\dagger.$$

we get convergence rates (in the Bregman distance) of order $\mathcal{O}(\delta)$. However, the benchmark results from the iterated Tikhonov-Morozov guarantee rates up to $\mathcal{O}(\delta^{2-\varepsilon})$ (under suitable source-conditions). These benchmarks results have not been obtained for the general ALM so far.

Our analysis shows that the dual variables in the ALM stay bounded, when the source condition is satisfied and the residual error decreases fast enough (actually these conditions are equivalent). We believe that conditions that guarantee faster convergence rates of the dual iterates is related to stronger source conditions, which in turn imply faster rates of the primal iterates.

Finally, we show that asymptotically the iterates of the ALM converge – when interpolated correctly – to solutions of the evolution equation

$$p' = g - Ku(t), \quad K^*p \in \partial J(u) \quad \text{and} \quad p(0) = p_0.$$

This equation system is a generalization of *Showalter's method* for regularization.

Acknowledgement

The work of O.S. has been supported by the Austrian Science Fund (FWF) within the national research networks *Industrial Geometry*, project 9203-N12, and *Photoacoustic Imaging in Biology and Medicine*, project S10505-N20. The work of K.F. was supported by the FWF within the national research networks Industrial Geometry, project 9203-N12 and the DFG-SNF Research Group FOR916 *Statistical Regularization*, Z-project.

References

[1] R. A. Adams. *Sobolev Spaces*, volume 65 of *Pure and Applied Mathematics*. Academic Press, New York - London, 1975.

[2] V. Barbu. *Nonlinear semigroups and differential equations in Banach spaces*. Editura Academiei Republicii Socialiste Romnia, Bucharest, 1976.

[3] H. Brézis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. Number 5 in North-Holland Mathematics Studies. North-Holland Publishing Co, Amsterdam-London, 1973.

[4] M. Burger, K. Frick, S. Osher, and O. Scherzer. Inverse total variation flow. *Multiscale Model. Simul.*, 6(2):365–395 (electronic), 2007.

[5] M. Burger, S. Osher, J. Xu, and G. Gilboa. Nonlinear inverse scale space methods for image restoration. In N. Paragios, O. D. Faugeras, T. Chan, and C. Schnörr, editors, *VLSM*, volume 3752 of *Lecture Notes in Computer Science*, pages 25–36. Springer, 2005.

[6] M. Burger, E. Resmerita, and L. He. Error estimation for bregman iterations and inverse scale space methods in image restoration. *Computing*, 81(2–3):109–135, 2007.

[7] T. F. Chan and S. Esedoḡlu. Aspects of total variation regularized L^1 function approximation. *SIAM J. Appl. Math.*, 65(5):1817–1837, 2005.

[8] I. Ekeland and R. Temam. *Convex analysis and variational problems*, volume 1 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam-Oxford, 1976.

[9] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

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

[11] K. Frick. *The Augmented Lagrangian Method and Related Evolution Equations*. PhD thesis, University of Innsbruck, 2008.

[12] K. Frick and O. Scherzer. Convex inverse scale spaces. In F. Sgallari, A. Murli, and N. Paragios, editors, *SSVM*, volume 4485 of *Lecture Notes in Computer Science*, pages 313–325. Springer, 2007.

[13] C. W. Groetsch. *Stable approximate evaluation of unbounded operators*. Number 1894 in Lecture Notes in Mathematics. Springer-Verlag, Berlin, 2007.

[14] C. W. Groetsch and O. Scherzer. Non-stationary iterated Tikhonov-Morozov method and third-order differential equations for the evaluation of unbounded operators. *Math. Methods Appl. Sci.*, 23(15):1287–1300, 2000.

[15] O. Güler. On the convergence of the proximal point algorithm for convex minimization. *SIAM J. Control Optim.*, 29(2):403–419, 1991.

[16] M. R. Hestenes. Multiplier and gradient methods. *J. Optimization Theory Appl.*, 4:303–320, 1969.

[17] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer. A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Problems*, 23(3):987–1010, 2007.

[18] S. Hu and N. S. Papageorgiou. *Handbook of multivalued analysis. Vol. I*, volume 419 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, 1997. Theory.

[19] K. Ito and K. Kunisch. An active set strategy based on the augmented Lagrangian formulation for image restoration. *M2AN Math. Model. Numer. Anal.*, 33(1):1–21, 1999.

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

[21] V. A. Morozov. *Methods for solving incorrectly posed problems*. Springer-Verlag, New York, 1984. Translated from the Russian by A. B. Aries, Translation edited by Z. Nashed.

[22] M. Nashed, editor. *Generalized inverses and applications*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1976.

[23] R. H. Nochetto, G. Savaré, and C. Verdi. A posteriori error estimates for variable time-step discretizations of nonlinear evolution equations. *Comm. Pure Appl. Math.*, 53(5):525–589, 2000.

[24] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Model. Simul.*, 4(2):460–489 (electronic), 2005.

[25] M. J. D. Powell. A method for nonlinear constraints in minimization problems. In *Optimization (Sympos., Univ. Keele, Keele, 1968)*, pages 283–298. Academic Press, London, 1969.

[26] E. Resmerita. On total convexity, Bregman projections and stability in Banach spaces. *J. Convex Anal.*, 11(1):1–16, 2004.

[27] R. T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.*, 1(2):97–116, 1976.

[28] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational methods in imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, New York, 2009.

[29] S. Setzer. Split bregman algorithm, douglas-rachford splitting and frame shrinkage. In [33], pages 464–476, 2009.

[30] D. Showalter. Representation and computation of the pseudoinverse. *Proc. Amer. Math. Soc.*, 18:584–586, 1967.

[31] D. W. Showalter and A. Ben-Israel. Representation and computation of the generalized inverse of a bounded linear operator between Hilbert spaces. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8)*, 48:184–194, 1970.

[32] E. Tadmor, S. Nezzar, and L. Vese. A multiscale image representation using hierarchical $(\mathrm{BV}, \mathrm{L}^2)$ decompositions. *Multiscale Model Simul.*, 2(4):554–579, 2004.

[33] X. Tai, K. Mørken, M. Lysaker, and K.-A. Lie, editors. *Scale Space and Variational Methods in Computer Vision*, volume 5567 of *Lecture Notes in Computer Science*, Berlin/Heidelberg, 2009. Springer. Second International Conference, SSVM 2009, Voss, Norway, June 1-5, 2009. Proceedings.

[34] D. Werner. *Funktionalanalysis*. Springer Verlag, Berlin, 2002.

[35] W. P. Ziemer. *Weakly differentiable functions*. Springer Verlag, New York, 1989.