DFG-SNF Research Group FOR916

Statistical Regularization and Qualitative Constraints

Klaus Frick

Dirk A. Lorenz

Elena Resmerita

Morozov's principle for the augmented Lagrangian method applied to linear inverse problems

Preprint FOR916 10-33

Updated Version (18-03-2011)

Preprint-Series of the Research Group FOR916

FOR916/10-33

October 2010

MOROZOV'S PRINCIPLE FOR THE AUGMENTED LAGRANGIAN METHOD APPLIED TO LINEAR INVERSE PROBLEMS

KLAUS FRICK[†], DIRK A. LORENZ[‡], AND ELENA RESMERITA[§]

Abstract. The Augmented Lagrangian Method as an approach for regularizing inverse problems received much attention recently, e.g. under the name Bregman iteration in imaging. This work shows convergence (rates) for this method when Morozov's discrepancy principle is chosen as a stopping rule. Moreover, error estimates for the involved sequence of subgradients are pointed out.

The paper studies implications of these results for particular examples motivated by applications in imaging. These include the total variation regularization as well as ℓ^q penalties with $q \in [1, 2]$. It is shown that Morozov's principle implies convergence (rates) for the iterates with respect to the metric of strict convergence and the ℓ^q-norm, respectively.

AMS subject classifications. 65J22, 46N10, 49M05

Key words. Augmented Lagrangian Method, Bregman iteration, Morozov's discrepancy principle, regularization

1. Introduction.

A classical problem in optimization is the solution of

$$J(u) \to \min \qquad \text{subject to} \qquad Ku = g\,, \tag{1.1}$$

where $J : H_1 \to \mathbb{R} \cup \{\infty\}$ is a convex functional and $K : H_1 \to H_2$ is a linear and bounded operator between Hilbert spaces H_1 and H_2. Solutions of problem (1.1) are called *J-minimizing solutions* of the equation $Ku = g$.

Of particular interest are ill-posed equations, that is, when the solution of $Ku = g$ does not depend continuously on the data g (as it is e.g. the case if K has non-closed range). This becomes distinctly delicate if the data g are not available precisely but only noise-affected observations g^δ for which we assume that we have the additional information $\left\|g^\delta - g\right\| \le \delta$.

It is a natural question to ask: "When does a solution algorithm for the optimization problem (1.1) applied to perturbed data g^δ instead of g, constitute a *regularization method* for the ill-posed equation $Ku = g$?" In [12, 20] an affirmative answer was given in a general setting for the *Augmented Lagrangian Method* (ALM), which in the context of quadratic regularization yields the *iterative Tikhonov method* (see, e.g. [17]), while the total variation setting is known as the *Bregman iteration* (see [20]). The ALM was introduced simultaneously by Hestenes [16] and Powell [21]—under the name *method of multipliers*—as an iterative solution method for (1.1) and reads as follows:

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

$$u_n^\delta \in \operatorname*{argmin}_{u \in H_1} \left(\frac{\tau_n}{2} \left\|Ku - g^\delta\right\|^2 + J(u) - \left\langle p_{n-1}^\delta, Ku - g^\delta \right\rangle \right) \qquad \textit{and} \tag{1.2a}$$

$$p_n^\delta = p_{n-1}^\delta + \tau_n (g^\delta - K u_n^\delta). \tag{1.2b}$$

[†]Institute for Mathematical Stochastics, University of Göttingen, Goldschmidtstraße 7 D-37077 Göttingen, Germany, `frick@math.uni-goettingen.de`

[‡]Institute for Analysis and Algebra, TU Braunschweig, D-38092 Braunschweig, Germany, `d.lorenz@tu-braunschweig.de`

[§]Industrial Mathematics Institute, Johannes Kepler University Linz, Altenbergerstrae 69, A-4040 Linz, Austria, `elena.resmerita@jku.at`

The name *Augmented Lagrangian* stems from the fact that the functional

$$\mathcal{L}(u,p) = J(u) - \langle p, Ku - g^\delta \rangle$$

is the Lagrangian for (1.1) and the additional term $\frac{\tau_n}{2} \left\| Ku - g^\delta \right\|^2$ is an *augmentation* of $\mathcal{L}$ that fosters the fulfillment of the constraint. Hence, in the limit, the augmentation term is supposed to vanish and the variables p_n^δ shall tend to a Lagrange multiplier for the problem (1.1).

It is well known that the Karush-Kuhn-Tucker conditions are necessary and sufficient regularity conditions for the solutions of (1.1), which guarantee existence of a saddle point of $\mathcal{L}$. Thus, if there exists $u^\dagger \in H_1$ and $p^\dagger \in H_2$ such that

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

then, $\mathcal{L}(u^\dagger, p) \leq \mathcal{L}(u^\dagger, p^\dagger) \leq \mathcal{L}(u, p^\dagger)$. It was pointed out in [4] that this coincides with the standard source condition in regularization theory.

As in [12], we will consider the ALM as a regularization method, that is, for stably computing approximations of solutions of (1.1) from perturbed data g^δ. Note that iterative methods (algorithms) for ill-posed problems behave as follows: The iterates start approximating a solution better and better, up to some iteration index when the contrary happens, due to error accumulation. Therefore, one needs to stop the algorithm at the right time in order to obtain stable approximations of the solution. With $\mathcal{R}_n : H_2 \to H_1$ and $\mathcal{R}_n^* : H_2 \to H_2$ we denote the operators defined by

$$\mathcal{R}_n(g^\delta) := u_n^\delta \;\text{ and }\; \mathcal{R}_n^*(g^\delta) = p_n^\delta, \text{ respectively.}$$

The paper [12] came up with a characterization of stopping index choice rules $\Gamma : (0,\infty) \times H_2 \to \mathbb{N}$ such that for each solution $u^\dagger$ of (1.1)

$$\mathcal{R}_{\Gamma(\delta_k, g_k)}(g_k) \to u^\dagger \quad \text{as} \quad \|g - g_k\| =: \delta_k \to 0$$

in an appropriate sense. Under a standard source condition, it showed also convergence rates for a class of stopping rules $\Gamma(\delta, y^\delta)$ for which $\Gamma(\delta, y^\delta) \to \infty$, as $\delta \to 0$.

We pursue further that study and focus on a widely used rule in applications, namely Morozov's discrepancy principle. Note that the complex challenge of choosing a right regularization parameter when dealing with stabilization methods for improperly posed problems is frequently approached via Morozov's rule due to its natural heuristic motivation. Namely, this rule selects a parameter by comparing the residual $\left\| Ku_n^\delta - g^\delta \right\|$ with the presumably known noise level δ—see, e.g. [11, Ch. 4]. We show that this rule does belong to the class of stopping rules studied in [12] and thus, obtain convergence rates for the primal iterates. Unlike [12], we investigate also the degenerate case of the discrepancy principle, that is when $\{\Gamma(\delta, g^\delta)\}$ has finite accumulation points. The analysis is non-trivial and provides a complete picture of the ALM behaviour when the discrepancy principle is utilized.

Moreover, we analyze in more detail two choices for J that are especially appealing for inverse problems occurring in *imaging*:

i) *Total-variation regularization* (cf. [4, 5, 24]). Let $H_1 = \mathrm{L}^2(\Omega)$ for a bounded domain $\Omega \subset \mathbb{R}^2$ and let $|\mathrm{D}u|(\Omega)$ denote the total-variation of the (measure-valued) distributional derivative of u. Then we consider the function

$$J(u) = \begin{cases} |\mathrm{D}u|(\Omega) & \text{if } u \in \mathrm{BV}(\Omega) \\ +\infty & \text{else.} \end{cases} \tag{1.3}$$

ii) *Sparse regularization* (cf. [9, 13, 18]). Let $H_1 = \ell^2$, $1 \le q \le 2$ and

$$J(u) = \begin{cases} \sum_{k\in\mathbb{N}} |u_k|^q & \text{if } \; u \in \ell^q \\ +\infty & \text{otherwise.} \end{cases} \tag{1.4}$$

Our results on these particular cases are complementary to the literature on Bregman iteration convergence. Thus, regarding ALM with total variation to which Morozov rule is employed, we show convergence rates for the approximations with respect to the strict convergence metric that is relevant in the BV space—see, for comparison [20] and [5] which establish weak* convergence and convergence rates with respect to Bregman distances, respectively. As of ALM combined with Morozov rule for sparsity problems, we derive convergence (rate) with respect to the ℓ^q-norm, $q \in [1, 2]$. Another novelty of our work regards error bounds for the divergence of the subgradients of J at the approximations and at the solution. Here divergence is measured by means of the Bregman distance with respect to the Fenchel conjugate J^*. This is relevant, e.g., for reconstruction of sparse signals. In this case, norm convergence rates are established for the subgradient of the ℓ^q-norm when $q \in (1, 2]$.

This work is organized as follows. Section 2 presents the main notions and notation, while Section 3 proposes several error estimates which extend related results in [12]. For instance, upper bounds for the Bregman distance between the subgradients of the objective functional J in (1.1) corresponding to the iterates and the solution, respectively, are obtained. We will utilize these estimates in Section 4 in order to show that the ALM together with Morozov's discrepancy principle lead to stable approximations for the operator equation both in the nondegenerate and degenerate cases. The results are applied for the total variation setting in Section 5, by underlying strict convergence (rates) for the primal variables. Section 6 summarizes the knowledge on the ALM for the sparsity regularization setting, i.e. convergence rates for the primal variables with respect to the ℓ^q-norm ($1 \le q \le 2$), as well as for the subgradients of these variables with respect to Bregman distances ($1 \le q \le 2$) and to dual norms ($1 < q \le 2$).

2. Preliminaries.

2.1. Basic Assumptions. Throughout this paper we will assume that H_1 and H_2 are separable Hilbert spaces with inner products $\langle \cdot, \cdot \rangle$ and norms $\|\cdot\|$ (not further specified since the meaning is always clear from the context). We will frequently make use of *Young's inequality*, which states that for all $u, v \in H_1$ and $\gamma > 0$ one has that

$$|\langle u, v \rangle| \le \frac{1}{2\gamma} \|u\|^2 + \frac{\gamma}{2} \|v\|^2 .$$

We assume further that $K : H_1 \to H_2$ is a linear and bounded operator and that $J : H_1 \to \overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ is convex, lower semi-continuous (l.s.c.) and proper, that is, the domain $D(J) = \{u \in H_1 \;:\; J(u) < \infty\}$ is non-empty. In order to guarantee that J-minimizing solutions of $Ku = g$ exist and that Algorithm 1 is well defined, we need to impose additional restrictions (cf. [12, Lem. 3.1]):

Assumption 1. *The sub-level sets*

$$\Lambda(c) = \left\{ u \in H_1 \;:\; \|Ku\|^2 + J(u) \le c \right\}$$

are weakly sequentially pre-compact, i.e. every sequence $\{u_n\}_{n\in\mathbb{N}}$ *contained in a sub-level set* $\Lambda(c)$ *has a weakly convergent subsequence.*

Moreover, we will assume that $\{\tau_n\}_{n\in\mathbb{N}}$ in Algorithm 1 is a fixed sequence of positive regularization parameters (which can be considered as step-sizes). We will make use of the quantity

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

The case of constant parameter $\tau_n = \tau$ is known as *stationary augmented Lagrangian method* and leads to $t_n = n\tau$. We will only require that the τ_n's do not decay too quickly and stay bounded, i.e.

$$\lim_{n\to\infty} t_n = +\infty \quad \text{and} \quad \sup_{k\in\mathbb{N}} \tau_k =: \bar{\tau} < \infty. \tag{2.1}$$

Finally, we will assume that $g \in H_2$ is an *attainable* element, that is, there exists a $u \in D(J)$ such that $Ku = g$. By $g^\delta \in H_2$ we always denote a perturbed version of g satisfying $\left\|g^\delta - g\right\| \leq \delta$ and then the sequences of primal and dual iterates of Algorithm 1 are encoded by $\left\{u_n^\delta\right\}_{n\in\mathbb{N}}$ and $\left\{p_n^\delta\right\}_{n\in\mathbb{N}}$ respectively.

In our convergence analysis we will make use of a sequence $\{g_k\}_{k\in\mathbb{N}} \in H_2$ that approach g as $k \to \infty$. To be more precise, we assume that $\{\delta_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^+$ is such that

$$\|g - g_k\| \leq \delta_k \to 0^+, \quad \text{as } k \to \infty.$$

In this case we will use the more explicit notation $\mathcal{R}_n(g_k)$ (and $\mathcal{R}_n^*(g_k)$) for the n-th primal (and dual) iterate of Algorithm 1.

2.2. Convex Analysis Tools. In the course of this paper we will frequently use some tools from convex analysis. A standard reference in this respect is [10].

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

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

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

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

According to [10, Ch. I Cor. 5.1], the set $\mathrm{Gr}(\partial J)$ is sequentially closed with respect to the weak-strong topology on $H_1 \times H_1$. That is, if the sequence $\{(u_n, \xi_n)\}_{n\in\mathbb{N}}$ of elements in $\mathrm{Gr}(\partial J)$ satisfies that u_n converges weakly to u and ξ_n converges strongly to ξ, then $(u,\xi) \in \mathrm{Gr}(\partial J)$.

The functional $J^* : H_1 \to \overline{\mathbb{R}}$ denotes the *Legendre-Fenchel transform* (or the dual functional) of J, which is defined by

$$J^*(v) := \sup_{u\in H_1} (\langle v, u\rangle - J(u)).$$

The functional J^* is the pointwise supremum of affine functions and hence it is convex, l.s.c. and proper [10, Ch. I, Prop. 3.1]. Moreover, one has [10, Ch. I, Cor. 5.2.]

$$v \in \partial J(u) \Leftrightarrow u \in \partial J^*(v).$$

Furthermore, it follows from the definition of the subgradient that

$$u \in \partial J^*(K^* p) \Rightarrow Ku \in \partial (J^* \circ K^*)(p).$$

For $u \in D(\partial J)$ and $v \in D(J)$, the *Bregman distance* of J between u and v with respect to $\xi \in \partial J(u)$ is defined by

$$D_J^\xi(v,u) = J(v) - J(u) - \langle \xi, v - u \rangle .$$

We will skip the superscript ξ, if the choice of the subgradient is obvious. If additionally $v \in D(\partial J)$ and $\eta \in \partial J(v)$, we further define the *symmetric* Bregman distance

$$D_J^{\mathrm{sym}}(v,u) = D_J(v,u) + D_J(u,v) = \langle \eta - \xi, v - u \rangle .$$

Note that the convexity of J implies that D_J and D_J^{sym} are always non-negative.

2.3. Source Condition. It is well known, that regularization methods for the reconstruction of a solution $u^\dagger$ of (1.1) in general converge arbitrarily slow, unless further regularity is imposed on $u^\dagger$ [11, Ch. 3, Prop. 3.11]. In the general setup presented in this paper, this is usually done in terms of the standard *source condition* [4], that is, there exists an element $p^\dagger \in H_2$ (the source element) such that

$$K^* p^\dagger \in \partial J(u^\dagger). \tag{2.2}$$

3. Extensions of previous error estimates results. In this section we derive extended error estimates as compared to [12, Thm. 6.2] that will place us in the position to find quantiative error bounds when Morozov's stopping rule is applied (cf. Section 4). However, these estimates are not solely a prerequesite but are interesting in itselves: They potentially can be combined with further stopping critera and they allow for convergence rates of the sequence $\left\{K^* p_n^\delta\right\}_{n\in\mathbb{N}}$ in the Bregman-distance associated with the Fenchel conjugate J^*. We start with some necessary background.

As it was first observed by Rockafellar in [23] the dual variables in the ALM can be characterized by the *proximal point method*. This characterization will play a central role in the following convergence analysis. In the current context, defining $G : H_2 \times H_2 \to \mathbb{R}$ by

$$G(p,g) = J^*(K^* p) - \langle p, g \rangle , \tag{3.1}$$

it holds that (cf. [12, Prop. 4.2])

$$p_n^\delta = \operatorname*{argmin}_{p \in H_2} \left(\frac{1}{2} \left\| p - p_{n-1}^\delta \right\|^2 + \tau_n G(p, g^\delta) \right) . \tag{3.2}$$

The basis of our analysis below is the following estimate on the iterates in (3.2) which was established by Güler in [15, Lem. 2.2]:

$$G(p_n^\delta, g^\delta) - G(p, g^\delta) \le \frac{\left\| p - p_0^\delta \right\|^2}{2t_n} - \frac{\left\| p - p_n^\delta \right\|^2}{2t_n} - \frac{t_n \left\| p_n^\delta - p_{n-1}^\delta \right\|^2}{2\tau_n^2}, \tag{3.3}$$

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

Next we prove error estimates for the primal and dual iterates of the ALM in case that the source condition (2.2) holds.

Theorem 3.1. *Assume that $u^\dagger$ is a J-minimizing solution of $Ku = g$ which satisfies the source condition* (2.2) *with source element $p^\dagger \in H_2$.*

(I) For any $\gamma > 0$

$$D_{J^*}^{u^\dagger}(K^*p_n^\delta, K^*p^\dagger) + \frac{t_n}{4}\left\|Ku_n^\delta - g\right\|^2 + \frac{\gamma-1}{2\gamma t_n}\left\|p_n^\delta - p^\dagger\right\|^2 \leq \frac{\left\|p^\dagger - p_0^\delta\right\|^2}{2t_n} + \frac{(1+\gamma)t_n\delta^2}{2}. \quad (3.4)$$

(II) For any $0 \leq \alpha < 1/2$,

$$D_{J^*}^{u^\dagger}(K^*p_n^\delta, K^*p^\dagger) + \frac{t_n(1-\alpha)}{4}\left\|Ku_n^\delta - g\right\|^2 + \alpha D_J^{sym}(u_n^\delta, u^\dagger) \leq \frac{1-\alpha}{1-2\alpha}\delta^2 t_n + \frac{\left\|p^\dagger - p_0^\delta\right\|^2}{2t_n}. \quad (3.5)$$

(III) It holds that

$$D_J^{sym}(u_n^\delta, u^\dagger) \leq \left\|Ku_n^\delta - g\right\|\left(\delta t_n + \sqrt{\delta^2 t_n^2 + \left\|p_0^\delta - p^\dagger\right\|^2}\right).$$

Proof. (I) Since $u^\dagger$ satisfies the source condition, we have that $K^*p^\dagger \in \partial J(u^\dagger)$ which is equivalent to $u^\dagger \in \partial J^*(K^*p^\dagger)$. This leads to

$$\begin{aligned} G(p_n^\delta, g^\delta) - G(p^\dagger, g^\delta) &= G(p_n^\delta, g) - G(p^\dagger, g) + \left\langle p_n^\delta - p^\dagger, g - g^\delta\right\rangle \\ &= J^*(K^*p_n^\delta) - J^*(K^*p^\dagger) - \left\langle p_n^\delta - p^\dagger, g\right\rangle + \left\langle p_n^\delta - p^\dagger, g - g^\delta\right\rangle \\ &= J^*(K^*p_n^\delta) - J^*(K^*p^\dagger) - \left\langle K^*p_n^\delta - K^*p^\dagger, u^\dagger\right\rangle + \left\langle p_n^\delta - p^\dagger, g - g^\delta\right\rangle \\ &= D_{J^*}^{u^\dagger}(K^*p_n^\delta, K^*p^\dagger) + \left\langle p_n^\delta - p^\dagger, g - g^\delta\right\rangle. \end{aligned}$$

Therefore, the last inequality together with (3.3) and Young's inequality gives for an arbitrary $\gamma > 0$

$$\begin{aligned} D_{J^*}^{u^\dagger}(K^*p_n^\delta, K^*p^\dagger) &= G(p_n^\delta, g^\delta) - G(p^\dagger, g^\delta) + \left\langle p_n^\delta - p^\dagger, g^\delta - g\right\rangle \\ &\leq \frac{\left\|p^\dagger - p_0^\delta\right\|^2}{2t_n} - \left(\frac{\gamma-1}{\gamma}\right)\frac{\left\|p^\dagger - p_n^\delta\right\|^2}{2t_n} - \frac{t_n\left\|p_n^\delta - p_{n-1}^\delta\right\|^2}{2\tau_n^2} + \frac{\gamma}{2}\delta^2 t_n. \quad (3.6) \end{aligned}$$

Using (1.2b) together with the inequality $\left\|Ku_n^\delta - g\right\|^2 \leq 2\left\|Ku_n^\delta - g^\delta\right\|^2 + 2\delta^2$ and the previous estimate show the assertion.

(II) From Young's inequality it follows that

$$\begin{aligned} \alpha D_J^{\mathrm{sym}}(u_n^\delta, u^\dagger) + \frac{t_n(1-\alpha)}{4}\left\|Ku_n^\delta - g\right\|^2 &= \alpha\left\langle p_n^\delta - p^\dagger, Ku_n^\delta - g\right\rangle + \frac{t_n(1-\alpha)}{4}\left\|Ku_n^\delta - g\right\|^2 \\ &\leq \frac{\alpha}{t_n}\left\|p_n^\delta - p^\dagger\right\|^2 + \frac{t_n}{4}\left\|Ku_n^\delta - g\right\|^2. \end{aligned}$$

Hence the first inequality follows from (I) with $\gamma = 1/(1-2\alpha)$, due to the fact that $\alpha < 1/2$.

In order to prove (III) we observe from (3.6) that for all $\gamma > 1$

$$\left\|p_n^\delta - p^\dagger\right\|^2 \leq \frac{\gamma}{\gamma-1}\left\|p_0^\delta - p^\dagger\right\|^2 + \frac{\gamma^2}{\gamma-1}\delta^2 t_n^2.$$

By elementary calculus, one can show that

$$\inf_{\gamma>1}\left(\frac{\gamma}{\gamma-1}a+\frac{\gamma^2}{\gamma-1}b\right)=\left(\sqrt{b}+\sqrt{a+b}\right)^2, \tag{3.7}$$

for any $a, b > 0$. Hence, letting $a = \left\|p_0^\delta - p^\dagger\right\|^2$ and $b = \delta^2 t_n^2$ leads to

$$\left\|p_n^\delta - p^\dagger\right\|^2 \le \left(\delta t_n + \sqrt{\delta^2 t_n^2 + \left\|p_0^\delta - p^\dagger\right\|^2}\right)^2.$$

Finally, the assertion follows from

$$D_J^{\mathrm{sym}}(u_n^\delta, u^\dagger) = \left\langle p_n^\delta - p^\dagger, K(u_n^\delta - u^\dagger)\right\rangle \le \left\|Ku_n^\delta - g\right\| \left\|p_n^\delta - p^\dagger\right\|.$$

∎

REMARK 3.2. Since $K^* p_n^\delta \in \partial J(u_n^\delta)$ and $K^* p^\dagger \in \partial J(u^\dagger)$ is equivalent to $u_n^\delta \in \partial J^*(K^* p_n^\delta)$ and $u^\dagger \in \partial J^*(p^\dagger)$ respectively, it follows that

$$\begin{aligned} D_{J^*}^{\mathrm{sym}}(K^*p_n^\delta, K^*p^\dagger) &= D_{J^*}^{u^\dagger}(K^*p_n^\delta, K^*p^\dagger) + D_{J^*}^{u_n^\delta}(K^*p^\dagger, K^*p_n^\delta) \\ &= \left\langle u_n^\delta - u^\dagger, K^*p_n^\delta - K^*p^\dagger\right\rangle \\ &= D_J^{\mathrm{sym}}(u_n^\delta, u^\dagger). \end{aligned}$$

Hence, all estimates for the primal variables $\left\{u_n^\delta\right\}_{n\in\mathbb{N}}$ automatically hold also for $\left\{K^* p_n^\delta\right\}_{n\in\mathbb{N}}$.

4. Morozov's discrepancy principle. In this section we analyze the discrepancy principle as an a posteriori stopping rule, proving that it guarantees stable approximations for the solution of the operator equation for which also error estimates can be derived. The study is done for both nondegenerate and degenerate cases, i.e, no matter how the stopping indices develop as the noise level decreases to zero.

We consider *Morozov's discrepancy principle*: Choose $\rho > 1$ and define

$$\Gamma(\delta, g^\delta) := \min\left\{n \in \mathbb{N} \;:\; \left\|Ku_n^\delta - g^\delta\right\| < \rho\delta\right\}. \tag{4.1}$$

That is, we take the first iterate u_n^δ for which the residual $\left\|Ku_n^\delta - g^\delta\right\|$ falls below a number which is a constant ρ times the noise level δ.

PROPOSITION 4.1. *The stopping rule* (4.1) *is well defined.*

Proof. It follows from [12, Cor. 5.2] that there exists a constant $C > 0$ such that

$$\frac{1}{2}\left\|Ku_n^\delta - g^\delta\right\|^2 \le \frac{C}{t_n} + \frac{\delta^2}{2}.$$

This implies that for all $\rho > 1$ there exists an index $n_0 \in \mathbb{N}$ for which $\left\|Ku_{n_0}^\delta - g^\delta\right\| < \rho\delta$. Thus, $\Gamma(\delta, g^\delta) < \infty$ is ensured. ∎

Our analysis is structured as follows: In Section 4.1 we study convergence of the primal iterates generated by the ALM with Morozov's rule as a stopping rule. Both nondegenerate and degenerate situations are considered. Note that the latter case has not been studied in [12] and thus, its analysis here is based on specific arguments. Moreover, the convergence (on subsequences) of the dual iterates also follows in that case. In Section 4.2 we derive convergence rates for the symmetric Bregman-distance between the primal iterates $\left\{u_n^\delta\right\}_{n\in\mathbb{N}}$ and J-minimizing solutions of $Ku = g$, under the hypothesis that the source condition holds, in case the stopping rule does not degenerate. Similar results are established for the degenerate case, the interesting part being that no source condition is needed in this respect.

4.1. Convergence analysis. In this section we prove convergence of the primal iterates in the ALM when Morozov's discrepancy principle (4.1) is employed as a stopping rule.

THEOREM 4.2. *Let Γ be chosen according to the stopping rule (4.1) and define*

$$\xi_k := K^*\mathcal{R}^*_{\Gamma(\delta_k,g_k)}(g_k) \in \partial J(\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k)).$$

(I) If

$$\lim_{k\to\infty} \Gamma(\delta_k, g_k) = \infty, \tag{4.2}$$

then the sequence $\left\{\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k)\right\}_{k\in\mathbb{N}}$ is bounded and each weak cluster point $u^\dagger$ is a J-minimizing solution of $Ku = g$. Additionally, it holds that

$$\lim_{k\to\infty} J(\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k)) = J(u^\dagger) \quad \text{and} \quad \lim_{k\to\infty} D_J^{\xi_k}(u^\dagger, \mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k)) = 0. \tag{4.3}$$

Moreover, the residuum satisfies the rate

$$\left\|K\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k) - g\right\| = \mathcal{O}(t^{-1/2}_{\Gamma(\delta_k,g_k)}). \tag{4.4}$$

(II) If (4.2) does not hold, then for each finite accumulation point $N \in \mathbb{N}$ of the stopping index sequence $\Gamma(\delta_k, g_k)$ one has

i) The set $\left\{p_N^\delta\right\}_{\delta>0}$ is bounded and each of its weak cluster points is a minimizer of $G(\cdot, g)$.

ii) The set $\left\{u_N^\delta\right\}_{\delta>0}$ is bounded and each of its weak cluster points is a J-minimizing solution of $Ku = g$.

Proof. (I) According to [12, Thm. 5.3], it is sufficient to show that $\Gamma(\delta_k, g_k)$ satisfies

$$\lim_{k\to\infty} t_{\Gamma(\delta_k,g_k)} = +\infty \quad \text{and} \quad \lim_{k\to\infty} \delta_k^2 t_{\Gamma(\delta_k,g_k)} = 0 \tag{4.5}$$

In order to show this, let $g^\delta \in H_2$ and set $\delta := \left\|g - g^\delta\right\|$ as well as $n_* = \Gamma(\delta, g^\delta) - 1$. Then, it follows from (4.1) that

$$\left\|Ku^\delta_{n^*} - g^\delta\right\| \geq \rho\delta.$$

This together with (3.3) yields

$$\frac{\left\|p - p^\delta_{n_*}\right\|^2}{2t_{n_*}} + \frac{\rho\delta^2 t_{n_*}}{2} \leq G(p, g^\delta) - G(p^\delta_{n_*}, g^\delta) + \frac{\left\|p - p^\delta_0\right\|^2}{2t_{n_*}}$$

for all $p \in H_2$ (recall that $\left\|Ku^\delta_n - g^\delta\right\| = \tau_n^{-1}\left\|p^\delta_n - p^\delta_{n-1}\right\|$ by (1.2b)). From the definition of G it follows that $G(p, g^\delta) - G(p^\delta_{n_*}, g^\delta) = G(p, g) - G(p^\delta_{n_*}, g) + \left\langle p - p^\delta_{n_*}, g - g^\delta\right\rangle$. After applying Young's inequality to the inner product we get, for every $p \in H_2$ and $\eta > 0$,

$$\frac{\left\|p - p^\delta_{n_*}\right\|^2}{2t_{n_*}} + \frac{\rho\delta^2 t_{n_*}}{2} \leq G(p, g) - G(p^\delta_{n_*}, g) + \frac{\left\|p - p^\delta_{n_*}\right\|^2}{2\eta} + \frac{\eta\left\|g - g^\delta\right\|^2}{2} + \frac{\left\|p - p^\delta_0\right\|^2}{2t_{n_*}}.$$

Setting $\eta = t_{n_*}$ hence gives

$$\frac{(\rho - 1)\delta^2 t_{n_*}}{2} \leq G(p, g) - G(p^\delta_{n_*}, g) + \frac{\left\|p - p^\delta_0\right\|^2}{2t_{n_*}}. \tag{4.6}$$

Let $\varepsilon > 0$ and choose $p_\varepsilon \in H_2$ such that $G(p_\varepsilon, g) \leq \inf_{q \in H_2} G(q, g) + \varepsilon$ (note that, due to [12, Lem. 4.1], the right hand side is finite whenever g is attainable). This together with the estimate (4.6) shows

$$\frac{(\rho-1)\delta^2 t_{n_*}}{2} \leq \varepsilon + \frac{\left\|p_\varepsilon - p_0^\delta\right\|^2}{2 t_{n_*}}.$$

According to (2.1), the conditions $\tau_k \leq \bar{\tau}$ for all $k \in \mathbb{N}$, and $\lim_{k\to\infty} \Gamma(\delta_k, g_k) = \infty$ imply $\lim_{k\to\infty} t_{\Gamma(\delta_k, g_k)} = +\infty$. Hence, substituting g_k for g^δ, δ_k for δ, and $\Gamma(\delta_k, g_k) - 1$for n_* shows

$$\limsup_{k\to\infty} \delta_k^2 t_{\Gamma(\delta_k,g_k)} \leq \limsup_{k\to\infty} \left(\frac{2\varepsilon}{\rho - 1} + \frac{\left\|p_\varepsilon - p_0^\delta\right\|^2}{2 t_{\Gamma(\delta_k,g_k)-1}} + \delta_k^2 \bar{\tau} \right) = \frac{2\varepsilon}{\rho-1}.$$

Since ε is arbitrary, this proves the statement.

(II) For simplicity of notation, assume that

$$\limsup_{\delta\to 0^+} \Gamma(\delta, g^\delta) = N. \tag{4.7}$$

i) The definition of $\Gamma(\delta, g^\delta)$ in (4.1) and the monotonicity of the residual $\left\|K u_n^\delta - g^\delta\right\|$ (cf. [12, Cor. 3.3]) imply

$$\left\|K u_N^\delta - g^\delta\right\| \leq \rho\delta, \qquad \text{for all } \delta > 0. \tag{4.8}$$

In particular, this yields $\left\|K u_N^\delta - g\right\| \leq (\rho+1)\delta$.

It was shown in the proof of [12, Thm 5.3] (by using Güler's estimate (3.3) and Young's inequality) that

$$\left\|p - p_N^\delta\right\|^2 \leq 2\left\|p - p_0^\delta\right\|^2 + 4 t_N^2 \delta^2 + 4 t_N (G(p,g) - \inf_{q\in H_2} G(q,g)) \qquad \text{for all } p \in H_2.$$

Choosing an arbitrary p such that $G(p,g) < 0$ implies

$$\limsup_{\delta\to 0^+} \left\|p_N^\delta\right\| =: A < \infty.$$

Now, let $\{\delta_k\}_{k\in\mathbb{N}}$ be such that $\delta_k \to 0^+$ and that $p_N^{\delta_k} \rightharpoonup \hat{p} \in H_2$. Due to the dual characterization (3.2) and to the equality $p_{N-1}^{\delta_k} - p_N^{\delta_k} = \tau_N (K u_N^{\delta_k} - g^{\delta_k})$, it follows that $K u_N^{\delta_k} - g^{\delta_k} \in \partial G(\cdot, g^{\delta_k})(p_N^{\delta_k})$. Since $G(p,g) = G(p, g^{\delta_k}) + \langle p, g^{\delta_k} - g\rangle$ for all $p \in H_2$, one has

$$K u_N^{\delta_k} - g \in \partial G(\cdot, g)(p_N^{\delta_k}).$$

Recall that the graph of the subgradient of a convex and lower semi-continuous functional is weakly-strongly closed. Therefore, inequality (4.8) yields

$$0 = \lim_{k\to\infty} K u_N^{\delta_k} - g \in \partial G(\cdot, g)(w\text{-}\lim_{k\to\infty} p_N^{\delta_k}) = \partial G(\cdot, g)(\hat{p}).$$

ii) From the definition of u_N^δ in (1.2a) and the fact that $p_{N-1}^\delta - p_N^\delta = \tau_N (K u_N^\delta - g^\delta)$ it follows (for δ small enough)

$$\begin{aligned} \frac{\tau_N}{2}\left\|K u_N^\delta - g^\delta\right\| + J(u_N^\delta) &\leq \frac{\tau_N}{2}\delta^2 + J(u^\dagger) + \langle p_{N-1}^\delta, g - K u_N^\delta\rangle \\ &\leq \frac{\tau_N}{2}\delta^2 + J(u^\dagger) + A(\rho+1)\delta + \tau_N \rho(\rho+1)\delta^2. \end{aligned}$$

In other words, $J(u_N^\delta) - J(u^\dagger) = \mathcal{O}(\delta)$ as $\delta \to 0^+$. This together with (4.8) shows that $\sup_{\delta>0}\left\{J(u_N^\delta) + \left\|Ku_N^\delta\right\|\right\} < \infty$ and consequently, according to Assumption 1, that $\left\{u_N^\delta\right\}_{\delta>0}$ is weakly compact and hence bounded. Thus, the statement follows from (4.8) and the lower semi-continuity of J. ∎

4.2. Convergence rates. Now we derive a qualitative estimate for the Bregman distance between the primal variables in the ALM and solutions of (1.1) if the source condition is satisfied and if the Morozov stopping rule is applied. In particular, this analysis sheds some light on the role of ρ in (4.1). We distinguish between the nondegenerate and the degenerate case (note that the source condition is not required as an assumption in the degenerate case, being automatically satisfied).

THEOREM 4.3. *Let $\rho > 1$ and Γ be chosen as in* (4.1).

(I) If (4.2) *holds and there exists a solution $u^\dagger$ of* (1.1) *which verifies the source condition* (2.2) *with source element $p^\dagger$, then*

$$\left\|K\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k) - g\right\| = \mathcal{O}(t^{-1}_{\Gamma(\delta_k,g_k)}) \quad \text{and} \quad \left\|\mathcal{R}^*_{\Gamma(\delta_k,g_k)}(g_k)\right\| = \mathcal{O}(1) \tag{4.9}$$

as $k \to \infty$. Moreover,

$$D_J^{sym}(\mathcal{R}_{\Gamma(\delta,g^\delta)}(g^\delta), u^\dagger) \le \left(1 + \mathcal{O}\left(\sqrt{\delta}\right)\right)\frac{\rho(\sqrt{\rho}+1)}{\sqrt{\rho-1}}\left\|p_0^\delta - p^\dagger\right\|\delta. \tag{4.10}$$

(II) If (4.2) *does not hold, then all J-minimizing solutions of $Ku = g$ satisfy the source condition* (2.2). *If $N \in \mathbb{N}$ is an accumulation point of the stopping index sequence $\Gamma(\delta_k, g_k)$, then*

$$\left\|Ku_N^\delta - g\right\| < (\rho+1)\delta, \qquad D_J^{sym}(u_N^\delta, u^\dagger) = \mathcal{O}(\delta). \tag{4.11}$$

Proof. (I) According to [12, Thm. 6.3] a sufficient condition for (4.9) is

$$\sup_{k\in\mathbb{N}} \delta_k t_{\Gamma(\delta_k,g_k)} < \infty. \tag{4.12}$$

Since $u^\dagger$ satisfies the source condition with source element $p^\dagger$, it follows from [12, Prop. 6.1] that $G(p^\dagger, g) \le G(p, g)$ for all $p \in H_2$. Moreover, using $p^\dagger$ instead of p in (4.6) shows

$$\frac{(\rho-1)\delta^2 t_{n_*}}{2} \le \frac{\left\|p^\dagger - p_0^\delta\right\|^2}{2t_{n_*}}$$

or in other words

$$\delta t_{\Gamma(\delta,g^\delta)} \le \frac{\left\|p^\dagger - p_0^\delta\right\|}{\sqrt{\rho-1}} + \delta\bar{\tau}. \tag{4.13}$$

This yields (4.12). It remains to establish (4.10). From (4.13) it follows that

$$\begin{aligned}
\delta^2 t^2_{\Gamma(\delta,g^\delta)} + \left\|p^\dagger - p_0^\delta\right\|^2 &\le \frac{1}{\rho-1}\left\|p^\dagger - p_0^\delta\right\|^2 + \frac{2\delta\bar{\tau}}{\sqrt{\rho-1}}\left\|p^\dagger - p_0^\delta\right\| + \delta^2\bar{\tau}^2 + \left\|p^\dagger - p_0^\delta\right\|^2 \\
&= \frac{\rho}{\rho-1}\left\|p^\dagger - p_0^\delta\right\|^2 + \delta\left(\frac{2\bar{\tau}}{\sqrt{\rho-1}}\left\|p^\dagger - p_0^\delta\right\| + \delta\bar{\tau}^2\right) \\
&= \frac{\rho}{\rho-1}\left\|p^\dagger - p_0^\delta\right\|^2 + \mathcal{O}(\delta).
\end{aligned}$$

This together with (4.13) and the fact that $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for all $a, b > 0$ implies

$$\delta t_{\Gamma(\delta, g^\delta)} + \sqrt{\delta^2 t^2_{\Gamma(\delta, g^\delta)} + \left\|p^\dagger - p_0^\delta\right\|^2} \leq \frac{\sqrt{\rho}+1}{\sqrt{\rho}-1}\left\|p^\dagger - p_0^\delta\right\| + \mathcal{O}(\sqrt{\delta}).$$

Since by construction in (4.1)

$$\left\|K\mathcal{R}_{\Gamma(\delta, g^\delta)}(g^\delta) - g^\delta\right\| < \rho\delta,$$

the assertion follows from Theorem 3.1 (III).

(II) Let $p^\dagger$ be a minimizer of $G(\cdot, g)$, which exists according to Theorem 4.2 (II) i). This and the definition of $G(p, g)$ in (3.1) implies

$$G(p^\dagger, g) - G(p_N^\delta, g^\delta) \leq \delta \left\|p_N^\delta\right\|.$$

Moreover, we deduce from the optimality condition of (1.2a) that $K^* p_N^\delta \in \partial J(u_N^\delta)$, which in turn implies that $K u_N^\delta \in \partial(J^* \circ K^*)(p_N^\delta)$. Using the definition of the subgradient and some rearrangements give

$$G(p^\dagger, g) - G(p_N^\delta, g^\delta) \geq -\delta\left(\left\|p^\dagger\right\| + \left\|p_N^\delta\right\|\right).$$

Since $\left\{\left\|p_N^\delta\right\|\right\}_{\delta>0}$ is bounded according to Theorem (4.2) (II) ii), the previous two estimates result in

$$\lim_{\delta\to 0^+} J^*(K^* p_N^\delta) - \left\langle p_N^\delta, g^\delta\right\rangle = \lim_{\delta\to 0^+} G(p_N^\delta, g^\delta) = G(p^\dagger, g) = J^*(K^* p^\dagger) - \left\langle p^\dagger, g\right\rangle. \tag{4.14}$$

Using once more the relation $K^* p_N^\delta \in \partial J(u_N^\delta)$ shows that $J^*(K^* p_N^\delta) + J(u_N^\delta) = \left\langle K^* p_N^\delta, u_N^\delta\right\rangle$ and consequently

$$J^*(K^* p_N^\delta) - \left\langle p_N^\delta, g^\delta\right\rangle + J(u_N^\delta) = \left\langle K u_N^\delta - g^\delta, p_N^\delta\right\rangle.$$

Now, let $u^\dagger$ be a J-minimizing solution of $Ku = g$ which exists according to Theorem 4.2 (II) ii). Taking the limit $\delta \to 0^+$ in the previous equality, using (4.8), (4.14), as well as the boundedness of $\left\{p_N^\delta\right\}_{\delta>0}$ and the fact that $J(u_N^\delta) \to J(u^\dagger)$ result in

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

that is, $K^* p^\dagger \in \partial J(u^\dagger)$. This proves the first statement in (II). Based on this, the second statement follows from Theorem (4.2) (II) i) and Theorem 3.1 (III) together with the first inequality in (4.11). □

Remark 4.4. The function

$$f(\rho) := \frac{\rho(\sqrt{\rho}+1)}{\sqrt{\rho}-1}$$

which appears in the right hand side of (4.10) is minimal for $\rho^* \simeq 1.6404$ with $f(\rho^*) \simeq 4.6753$. Hence ρ^* might be a reasonable choice for ρ in the stopping rule (4.1).

Remark 4.5. a) In the nondegenerate case, it follows also that each weak cluster point of $\left\{\mathcal{R}^*_{\Gamma(\delta_k, g_k)}(g_k)\right\}_{k\in\mathbb{N}}$ is a minimizer of $G(\cdot, g)$ (see [12, Thm. 6.3]). From Schauder's Theorem and from $\overline{\operatorname{ran}(K)} = \ker(K^*)^\perp$ it follows that for each compact K with dense range, the adjoint operator K^* is compact and injective and hence

$$\lim_{k\to\infty} K^* \mathcal{R}^*_{\Gamma(\delta_k, g_k)}(g_k) = K^* \bar{p}$$

strongly, where $\bar{p}$ is a minimizer of $G(\cdot, g)$. If the condition on the range of K is not satisfied, then strong convergence hold on subsequences.

b) Regarding the degenerate case, $\{\Gamma(\delta, g^\delta)\}_{\delta>0}$ has finite accumulation points, without restricting generality, we can consider that this is a constant subsequence. This yields that for all δ sufficiently small, one has to stop the algorithm at the same iteration.

A degenerate case is discussed for the Landweber method for nonlinear equations in the book [11, p. 284]. It is shown there that $\lim_{\delta\to 0} u_N^\delta = u_N$ where u_N is the N-th iterate in the exact data case and is a solution of the operator equation as well. This means that in the exact data case the Landweber algorithm reaches the solution after N steps, with N being the stopping index in the noisy data case.

For the ALM analyzed here, we could not show that $\lim_{\delta\to 0} u_N^\delta = u_N$ where u_N is the N-th iterate in the exact data case because the implicit feature of the method makes the analysis more difficult. However, we could establish that the accumulation points of $\{u_N^\delta\}_{\delta>0}$ are J - minimizing solutions with additional smoothness, i.e., satisfying the source condition.

5. Iterative total variation regularization. The ALM method in the case of J being the total variation seminorm (1.3) is also known as Bregman iteration [20]. It was shown in [20] that Morozov's discrepancy principle yields weak* convergence in $\mathrm{BV}(\Omega)$ of the iterative method. The expected but missing convergence there was the one with respect to the total variation seminorm, in the sense

$$\lim_{k\to\infty} J(u_k) = J(u). \tag{5.1}$$

As a consequence of the analysis based on the augmented Lagrangian method tools, it became clear that this convergence does hold. Moreover, linear convergence rates with respect to the Bregman distance associated with the total variation seminorm were established in [5] first for the noise free case. In the noisy data case, an a posteriori stopping rule was analyzed in [20]:

$$n_*(\delta, g^\delta) = \max\left\{n \in \mathbb{N} : \|Ku_n^\delta - g^\delta\| \geq \rho\delta\right\}, \quad \rho > 1.$$

Although convergence was shown there for the net $\{u^\delta_{n_*(\delta,g^\delta)}\}$ as $\delta \to 0$, no convergence rate was obtained for it. This section aims to point out such a convergence rate. Note that the a posteriori rule (4.1) employed here relates to the above mentioned one by

$$\Gamma(\delta, g^\delta) = n_*(\delta, g^\delta) + 1.$$

Still, the question on how to quantify the weak* convergence is not answered. A possible answer could be given by taking into account that weak* convergence in $\mathrm{BV}(\Omega)$ together with convergence in the sense (5.1) is equivalent to so-called *strict convergence*. Thus, one can obtain convergence rates with respect to a related metric, as shown below. Recall [2, p. 125] that, by definition, $\{u_k\}_{k\in\mathbb{N}} \subset \mathrm{BV}(\Omega)$ converges *strictly* to u if it converges with respect to the metric

$$\tilde{d}(u, v) = \|u - v\|_{\mathrm{L}^1} + |J(u) - J(v)|. \tag{5.2}$$

PROPOSITION 5.1. *Let* Γ *be chosen according to the Morozov's rule* (4.1) *and assume that* $\lim_{k\to\infty} \Gamma(\delta_k, g_k) = \infty$. *Then, the sequence* $\left\{\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k)\right\}_{k\in\mathbb{N}}$ *satisfies*

(4.3) *and* (4.4). *Moreover, it has a subsequence which converges strictly to a J-minimizing solution of $Ku = g$.*

Proof. The first assertions result from Theorem 4.2 (I). Let further denote $u_k = \mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k)$. According to Theorem 4.2 (I), the sequence $\{u_k\}_{k\in\mathbb{N}}$ is bounded in $\mathrm{L}^2(\Omega)$ and $\sup_{k\in\mathbb{N}} J(u_k) < \infty$. Hence we find that

$$\sup_{k\in\mathbb{N}} \|u_k\|_{\mathrm{BV}} = \sup_{k\in\mathbb{N}} \|u_k\|_{\mathrm{L}^1} + J(u_k) < \infty.$$

Theorem 2.5 in [1] implies that $\{u_k\}_{k\in\mathbb{N}}$ is strongly L^1-compact and thus there is a subsequence, indexed by k', which converges to some u^* strongly in $\mathrm{L}^1(\Omega)$. Since each L^2-weak cluster point of $\{u_k\}_{k\in\mathbb{N}}$ is a J-minimizing solution of $Ku = g$ according to Theorem 4.2 (I), the same holds for u^*. Finally, it follows from (4.3) that $\tilde{d}(u_{k'}, u^*) \to 0$. □

Clearly, error estimates in terms of the L^1-norm are desirable, but not easy to derive. In order to show convergence rates for strict convergence of the iterates, we need to employ another metric, which appears naturally in the analysis, namely

$$d(u, v) = \|Ku - Kv\|_{\mathrm{L}^2} + |J(u) - J(v)|\,. \tag{5.3}$$

The following lemma points out the relation between the two metrics $\tilde{d}$ and d.

LEMMA 5.2. *Assume that $K : \mathrm{L}^1(\Omega) \to \mathrm{L}^2(\Omega)$ is continuous and can be extended by continuity to $\mathrm{L}^2(\Omega)$. Then, convergence of a sequence with respect to the metric $\tilde{d}$ defined by* (5.2) *implies convergence of the sequence with respect to the metric d defined by* (5.3). *If additionally the linear bounded operator $K : \mathrm{L}^2(\Omega) \to \mathrm{L}^2(\Omega)$ is injective, then the two metrics are equivalent.*

Proof. The first part follows immediately from $\|Ku\|_{\mathrm{L}^2} \le \|K\|\,\|u\|_{\mathrm{L}^1}$ for any $u \in \mathrm{L}^1(\Omega)$.

Assume now that $d(u_k, u) \to 0$ as $k \to \infty$ and that K is injective. Then, K in particular does not annihilate constant functions and it follows from [1, Lem. 4.1] that $u \mapsto \|Ku\|_{\mathrm{L}^2} + J(u)$ is BV-coercive. Hence boundedness of $\{\|Ku_k\|_{\mathrm{L}^2}\}_{k\in\mathbb{N}}$ and $\{J(u_k)\}_{k\in\mathbb{N}}$, which follows from $d(u_k, u) \to 0$, yields boundedness of $\{\|u_k\|_{\mathrm{BV}}\}_{k\in\mathbb{N}}$. Thus, there exists a subsequence $\{u_{k'}\}_{k'\in\mathbb{N}}$ which converges to some $v \in \mathrm{BV}(\Omega)$ strongly in $\mathrm{L}^1(\Omega)$ and weakly in $\mathrm{L}^2(\Omega)$ to v due to compact and bounded embedding respectively (cf. [1, Thm. 2.5]). These yield strong convergence of the subsequence in $\mathrm{L}^1(\Omega)$ to v, as well as weak convergence in $\mathrm{L}^2(\Omega)$ of $\{Ku_{k'}\}_{k'}$ to Kv. Since the weak limit is unique, it follows that $Ku = Kv$ and consequently, since K is injective, that $u = v$. Moreover, the entire sequence $\{u_k\}_{k\in\mathbb{N}}$ converges strongly in $\mathrm{L}^1(\Omega)$ to u, which completes the proof. □

Note that the continuity of the operator K from $\mathrm{L}^1(\Omega)$ into $\mathrm{L}^2(\Omega)$ is not necessary for proving the second part of the lemma.

Now we show the convergence rate in terms of the metric d:

PROPOSITION 5.3. *Let Γ be chosen according to rule* (4.1) *and assume that $\lim_{k\to\infty} \Gamma(\delta_k, g_k) = \infty$. If $u^\dagger$ is a J-minimizing solution of $Ku = g$ that satisfies the source condition* (2.2) *with source element $p^\dagger \in H_2$, then the following convergence rate holds:*

$$d(\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k), u^\dagger) = \|K\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k) - Ku^\dagger\| + |J(\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k)) - J(u^\dagger)| = \mathcal{O}(\delta_k).$$

Proof. From the definition of rule (4.1) it follows that $\|K\mathcal{R}_{\Gamma(\delta_k,g_k)}g_k - Ku^\dagger\| = \mathcal{O}(\delta_k)$. Hence in order to establish the assertion it remains to derive an error estimate for $|J(\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k)) - J(u^\dagger)|$. Since the symmetric Bregman distance is larger than the Bregman distance, one has

$$J(u^\dagger) - J(\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k)) \leq \\ \left\langle \mathcal{R}^*_{\Gamma(\delta_k,g_k)}(g_k), g - K\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k) \right\rangle + D_J^{\mathrm{sym}}(\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k), u^\dagger).$$

Using the Cauchy-Schwarz inequality and Theorem 4.3 (I) we see that

$$J(u^\dagger) - J(\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k)) = \mathcal{O}(\delta_k).$$

Similarly one can show

$$J(\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k)) - J(u^\dagger) = \mathcal{O}(\delta_k)$$

which ends the proof. ∎

6. Sparse regularization. In the case of sparse regularization, the convex functional (1.4) is considered with $1 \leq q \leq 2$ (see [9]). The aim of the functional J is to promote sparse solutions, i.e. solutions which have only a few (especially a finite number of) nonzero entries. Tikhonov regularization based on this regularization functional has been studied in great detail in [13, 18, 19]. The case $q = 1$ for the stationary augmented Lagrangian method has been treated in [5] also under the name Bregman iteration. There, the authors obtained convergence of the method for noisefree data for the Bregman distance and considered an a priori stopping rule for noisy data. In this section we also treat the case $q = 1$ and derive both an enhanced convergence rate for noisefree data in norm and also optimal convergence rates for noisy data with the a posteriori rule (4.1).

6.1. Convergence rates for $\delta \to 0$. We start with a result on convergence in the noisy data case which holds for all $q \in [1, 2]$. Fulfillment of a source condition is not needed here.

THEOREM 6.1. *Let $K : \ell^2 \to H_2$ be linear and bounded, $1 \leq q \leq 2$ and let J be defined by* (1.4). *Moreover, let the parameter choice Γ obey* (4.5). *Then the sequence $\{\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k)\}$ has a subsequence which converges strongly to a J-minimizing solution of $Ku = g$.*

Proof. By Theorem 4.2 (I), the sequence $\{\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k)\}$ is bounded in ℓ^2 and hence, has a subsequence which converges weakly in ℓ^2. Moreover, it follows from Theorem 4.2 (I) that $J(\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k)) \to J(u^\dagger)$. By [9, Lem. 4.3] this shows that $J(\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k))$ also converges strongly. ∎

Note that the entire sequence of iterates converges strongly to the unique J-minimizing solution of $Ku = g$ in the case $q \in (1, 2]$. By Theorem 4.2, we also conclude that ℓ^q-regularization combined with Morozov's discrepancy principle gives rise to a (subsequentially) convergent regularization method and, if additionally the source condition is fulfilled, leads to convergence rates in the sense of Bregman distances (cf. Theorem 4.3). Actually, in the latter case, we can strengthen the above result. More precisely, we can derive convergence rates with respect to the ℓ^q norm for $q \in [1, 2]$. The two cases $q \in (1, 2]$ and $q = 1$ have to be treated separately.

In the first case we take advantage of the differentiability and the high degree of convexity of the functional J to estimate even the distance between the subgradients appearing in the iterative process. The Fenchel conjugate of J is $J^*(\xi) = \frac{1}{r}\|\xi\|_{\ell^r}^r$ with $r = q/(q-1) > 2$ (see, e.g., [10, Prop. 4.2, p. 19].

The following result, which will be useful in the sequel, was pointed out in [22, Prop. 3.2]. We give here the proof for the sake of completeness.

LEMMA 6.2. *If $q \in (1,2]$ and J is given by (1.4), then for all $v \in \ell^q$ and $u \in \ell^{2(q-1)}$, with $\|v-u\|_{\ell^q}$ small enough there is a constant $c_q > 0$ depending on u such that*

$$D_J(v,u) \geq c_q \|v-u\|_{\ell^q}^2 . \tag{6.1}$$

Proof. Note that $D(\partial J) = \ell^{2(q-1)}$. In order to simplify the notation in the proof, we omit the subscript for the ℓ^q norm. The inequality is obvious if $q=2$. Let $q \in (1,2)$. If $u=0$, then $D_J(v,0) = \|v\|^q/q \geq \|v\|^2/q$ if $\|v\| \leq 1$, which yields (6.1). Now consider $u \in \ell^{2(q-1)}$, $u \neq 0$. [6, Lem. 1.4.8] implies that for all $v \in \ell^q$, $u \in \ell^{2(q-1)}$

$$D_J(v,u) \geq (t + \|u\|)^q - \|u\|^q - qt\|u\|^{q-1}, \tag{6.2}$$

where $t := \|v-u\|$. Let $\varphi(t) := (t+\|u\|)^q$ for t small enough. The Taylor expansion of φ around 0 yields existence of an $a_t \in (0,t)$ such that

$$\varphi(t) = \|u\|^q + qt\|u\|^{q-1} + \frac{q(q-1)t^2}{2}\|u\|^{q-2} + \frac{q(q-1)(q-2)t^3}{6}(a_t + \|u\|)^{q-3} .$$

This inequality and (6.2) imply

$$\begin{aligned} D_J(v,u) &\geq \varphi(t) - \|u\|^q - qt\|u\|^{q-1} \\ &= \frac{q(q-1)t^2}{2}\|u\|^{q-2} + \frac{q(q-1)(q-2)t^3}{6}(a_t + \|u\|)^{q-3} \\ &= \frac{q(q-1)t^2}{2}\|u\|^{q-2}\left[1 - \frac{(2-q)t}{3}\|u\|^{2-q}\left(a_t + \|u\|^{q-3}\right)\right]. \end{aligned}$$

Note that $a_t + \|u\| \geq \|u\|$ and $q-3<0$. Hence, $(a_t + \|u\|)^{q-3} \leq \|u\|^{q-3}$ and

$$D_J(v,u) \geq \frac{q(q-1)t^2}{2}\|u\|^{q-2}\left[1 - \frac{(2-q)t}{3}\|u\|^{-1}\right]. \tag{6.3}$$

Let $b \in (0,1)$ and take $t < \frac{3(1-b)\|u\|}{2-q}$. Then inequality (6.3) yields

$$D_J(v,u) \geq c_q t^2,$$

with $c_q = \frac{bq(q-1)}{2}\|u\|^{q-2}$. ∎

PROPOSITION 6.3. *Let $K : \ell^2 \to H_2$ be linear and bounded, J be defined by (1.4) with $1 < q \leq 2$ and $r = q/(q-1)$. Let Γ be the parameter choice according to Morozov's discrepancy principle (4.1). If the J-minimizing solution $u^\dagger$ of $Ku = g$ satisfies the source condition (2.2) with a source element $p^\dagger$, then the following convergence rates hold for k sufficiently large:*

$$\left\|\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k) - u^\dagger\right\|_{\ell^q} = \mathcal{O}(\sqrt{\delta_k}),$$

$$\left\|K^*\mathcal{R}^*_{\Gamma(\delta_k,g_k)}(g_k) - K^*p^\dagger\right\|_{\ell^r} = \mathcal{O}(\delta_k^{\frac{q-1}{q}}).$$

Proof. We apply inequality (6.1) for $v = \mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k)$ and $u = u^\dagger$ and obtain

$$D_J^{K^* p^\dagger}(\mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k), u^\dagger) \geq c_q \left\| \mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k) - u^\dagger \right\|_{\ell^q}^2 ,$$

for k sufficiently large. This and Theorem 4.3 (I) imply the first assertion.

In order to show the estimate for the subgradients, note that (see, e,g., [6, Lem. 1.4.10])

$$D_{J^*}(\xi_2, \xi_1) \geq c_r \left\| \xi_2 - \xi_1 \right\|_{\ell^r}^r$$

for any $\xi_1, \xi_2 \in \ell^r$for some positive constant c_r depending on $r \geq 2$. Consequently, it follows from Remark 3.2 that

$$\left\| K^* \mathcal{R}^*_{\Gamma(\delta_k,g_k)}(g_k) - K^* p^\dagger \right\|_{\ell^r} = \mathcal{O}(\delta_k^{\frac{q-1}{q}})$$

and thus completes the proof. ☐

Now we turn to the case of sparse regularization for $q = 1$. Here, one can derive improved convergence rates in case the solution $u^\dagger$ does not only fulfill the source condition but also is indeed sparse. To be more precise, we define for a given set $I \subset \mathbb{N}$ the projection P_I by

$$(P_I u)_k = \begin{cases} u_k, & k \in I \\ 0, & k \notin I. \end{cases}$$

and require the following

Assumption 2.

i) The solutions $u^\dagger$ of (1.1) *satisfy the source condition* (2.2) *with source element $p^\dagger$.*

ii) For $K^* p^\dagger = \xi$ and $I = \{k \mid |\xi_k| = 1\}$, one has that the quantity $\theta = \sup\{|\xi_k| \mid k \notin I\}$ is strictly smaller than one.

iii) The operator $KP_I : \ell^1 \to H$ is injective in the sense that $P_I u \neq P_I v$ implies $KP_I u \neq KP_I v$.

We start with the following lemma which can be traced back to [13] (see also [14]).

Lemma 6.4. *Assume that Assumption 2 is satisfied. Then, there exist constants $\beta_1, \beta_2 > 0$ such that*

$$J(u) - J(u^\dagger) \geq \beta_1 J(u - u^\dagger) - \beta_2 \left\| K(u - u^\dagger) \right\| .$$

Proof. Due to Assumption 2 iii) the operator KP_I is injective and hence, there exists c such that $\|KP_I u\| \geq c \|P_I u\|_{\ell^1}$ for all $u \in \ell^1$. Now we estimate

$$\begin{aligned} J(u - u^\dagger) &= \left\| u - u^\dagger \right\|_{\ell^1} = \left\| P_I(u - u^\dagger) \right\|_{\ell^1} + \|P_{I^\complement} u\|_{\ell^1} \\ &\leq \tfrac{1}{c} \left\| KP_I(u - u^\dagger) \right\| + \|P_{I^\complement} u\|_{\ell^1} \\ &\leq \tfrac{1}{c} \left\| K(u - u^\dagger) \right\| + (\|K\| + 1) \|P_{I^\complement} u\|_{\ell^1} . \end{aligned}$$

Since $u_k^\dagger = 0$ for $k \notin I$, Assumption 2 i) and ii) implies

$$\begin{aligned}\|P_{I^\complement} u\|_{\ell^1} &= \sum_{k\notin I} |u_k| \\ &\le \frac{1}{1-\theta}\Big(\sum_{k\notin I} |u_k| - \Big|u_k^\dagger\Big| - \xi_k(u_k - u_k^\dagger)\Big) \\ &\le \frac{1}{1-\theta}\Big(\sum_{k} |u_k| - \Big|u_k^\dagger\Big| - \xi_k(u_k - u_k^\dagger)\Big) \\ &= \frac{1}{1-\theta}\Big(J(u) - J(u^\dagger) - \big\langle \xi, u - u^\dagger\big\rangle\Big) \\ &\le \frac{1}{1-\theta}\Big(J(u) - J(u^\dagger) + \big\|p^\dagger\big\| \, \big\|K(u-u^\dagger)\big\|\Big).\end{aligned}$$

Combining both estimates gives

$$J(u-u^\dagger) \le \big(\tfrac{1}{c} + \big\|p^\dagger\big\| \frac{\|K\|+1}{1-\theta}\big) \big\|K(u-u^\dagger)\big\| + \frac{\|K\|+1}{1-\theta}(J(u) - J(u^\dagger))$$

which yields the assertion with

$$\beta_1 = \frac{1-\theta}{\|K\|+1}, \qquad \beta_2 = \frac{1-\theta}{(\|K\|+1)c} + \big\|p^\dagger\big\|.$$

∎

Remark 6.5. We remark on Assumption 2: Statement ii) is related to the notion of "strict sparsity pattern" in [3]. To get a practically relevant condition, one may replace this with the assumption that the range of K^* is contained in some ℓ^p with $p < \infty$ (since in this case the sequence ξ has to tend to zero). This also implies that I is finite. Alternatively one may also work with $K : \ell^2 \to H$ (which implies $K^* : H \to \ell^2$).

Assumption iii) is a restricted injectivity condition. Since one needs to know the set I to verify this in advance, one often uses the "finite basis injectivity property" (FBI property) from [3, 18] which states that KP_I is injective for all finite sets I. This condition can be checked in advance and hence, it seems more practical.

Now we treat the case of noisy data and show that the application of Morozov's discrepancy principle leads to optimal convergence rates.

Theorem 6.6. *Let $u^\dagger$ be a J-minimizing solution of $Ku = g$ and assume that Γ is the parameter choice according to Morozov's discrepancy principle* (4.1)*. Then, one has*

$$\big\|\mathcal{R}_{\Gamma(\delta_k,g_k)} g_k - u^\dagger\big\|_{\ell^1} = \mathcal{O}(\delta_k).$$

Proof. We estimate the symmetric distance from below using Lemma 6.4. To this end, set $u_k = \mathcal{R}_{\Gamma(\delta_k,g_k)}(g_k)$ and observe that

$$\begin{aligned} D_J^{\mathrm{sym}}(u_k, u^\dagger) &\ge D_J(u_k, u^\dagger) = J(u_k) - J(u^\dagger) + \big\langle K^* p^\dagger, u_k - u^\dagger \big\rangle \\ &\ge \beta_1 J(u_k - u^\dagger) - \beta_2 \|Ku_k - g\| - \big\langle p^\dagger, Ku_k - g\big\rangle.\end{aligned}$$

Rearranging and using the Cauchy-Schwartz inequality leads to

$$\beta_1 J(u_k - u^\dagger) \le D_J^{\mathrm{sym}}(u_k, u^\dagger) + \big(\beta_2 + \big\|p^\dagger\big\|\big) \|Ku_k - g\|.$$

From the definition of Morozov's discrepancy principle (4.1) and Theorem 4.3 (I) we finally conclude the proof. ∎

6.2. Convergence rate for $n \to \infty$ in the noisefree case. As another consequence of our analysis of the ALM we can prove convergence rates of the ALM iteration with noisefree data for $q = 1$ which are superior to previous results. To this end, let u_n and p_n be the primal and dual iterates generated by the Algorithm 1 with g^δ is replaced by g.

PROPOSITION 6.7. *Let J be according to* (1.4) *with $q = 1$, $u^\dagger$ be a J-minimizing solution of $Ku = g$ and $p_0 = 0$. Then there exists a constant $C > 0$ such that*

$$\left\| u_n - u^\dagger \right\|_{\ell^1} \leq C t_n^{-1}.$$

Proof. Since $K^* p_n \in \partial J(u_n)$, one has

$$J(u_n) - J(u^\dagger) \leq \left\langle K^* p_n, u_n - u^\dagger \right\rangle = \langle p_n, K u_n - g \rangle \leq \|p_n\| \, \|K u_n - g\| \, .$$

Now we use Lemma 6.4 to obtain

$$\beta_1 J(u_n - u^\dagger) \leq J(u_n) - J(u^\dagger) + \beta_2 \|K u_n - g\| \leq (\|p_n\| + \beta_2) \|K u_n - g\| \, .$$

Theorem 3.1 (with $\delta = 0$) finally gives

$$J(u_n - u^\dagger) \leq \frac{(\gamma + \beta_2) \left\| p^\dagger \right\|}{\beta_1 t_n}. \qquad \square$$

6.3. Implications for Compressed Sensing. Finally we remark on the relation of our results to the theory of compressed sensing: Linear convergence rates for the variational regularization with ℓ^1-norm has been shown in [13, 14] under a source condition and some assumptions on the operator K. A similar result has been proven (see [7]) in the finite dimensional setting of compressed sensing, by using the restricted isometry property condition. In this setting, [14, part of Prop. 5.3 and Thm. 4.7] established that the s-restricted isometry property for K implies that an s-sparse solution $u^\dagger$ satisfies the source condition (2.2) and KP_I is injective, with I given by Lemma 6.4. As an immediate consequence. we can deduce that a linear convergence rate s-sparse solutions for the Bregman iteration both in the noisefree case and in the noisy case with discrepancy principle holds under the s-restricted isometry property.

7. Conclusion. In this work we showed that Morozov's discrepancy principle (4.1) applied to the Augmented Lagrangian Method (ALM) leads to a regularization method for linear inverse problems $Ku = g$. This gives a theoretical justification for the observation that the discrepancy principle provides useful results in practical situations.

We used a dual characterization of the ALM in order to derive explicit error bounds for the Bregman distance between the iterates and a true J-minimizing solution $u^\dagger$ of $Ku = g$, if $u^\dagger$ satisfies the source condition

$$K^* p^\dagger \in \partial J(u^\dagger)$$

for a source element $p^\dagger$. In this case, also error bounds for the Bregman distance (with respect to J^*) between the dual iterates in the ALM and $p^\dagger$ were obtained. We also showed that a *sufficient condition* for the source condition to hold is the existence of finite accumulation points in the sequence of stopping indices chosen by the discrepancy principle.

In the case of total variation regularization we were able to show that the ALM converges *strictly* in $\mathrm{BV}(\Omega)$ and established convergence rates with respect to an equivalent metric.

In the case of sparse regularization, more precisely when J coincides with the ℓ^q-norm ($q \in [1,2]$) we derived $\sqrt{\delta}$-rates in the ℓ^q-norm for $q > 1$ and linear rates for the particular interesting case of ℓ^1 (under suitable regularity conditions on $u^\dagger$). The sequence of dual iterates in the latter case carries important information on the support of the solution. The conjugate function J^* of the ℓ^1-norm, however, degenerates to an indicator function. As a consequence, the general estimates for the dual variables do not reveal much insight in their convergence behavior. It is still an open issue whether one can obtain more relevant estimates for the dual variables.

Acknowledgments. The authors thank Martin Burger (University of Münster) for stimulating discussions on the total variation section. The referees' constructive comments, which led to an improved presentation of the paper, are gratefully acknowledged. K.F. is supported by the DFG-SNF Research Group FOR916 *Statistical Regularization* (Z-Project). E.R. acknowledges the support by Austrian Science Fund, project FWF V82-N118 (Elise Richter fellowship). D.L. acknowledges support by the DFG under grant LO 1436/2-1 within the DFG priority program SPP 1324.

References.

[1] Robert Acar and Curtis Vogel, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems, 10 (1994), pp. 1217–1229.

[2] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 2000.

[3] Kristian Bredies and Dirk A. Lorenz, *Linear convergence of iterative soft-thresholding*, Journal of Fourier Analysis and Applications, 14 (2008), pp. 813–837.

[4] Martin Burger and Stanley J. Osher, *Convergence rates of convex variational regularization*, Inverse Problems, 20 (2004), pp. 1411–1420.

[5] Martin Burger, Elena Resmerita, and Lin He, *Error estimation for Bregman iterations and inverse scale space methods in image restoration*, Computing, 81 (2007), pp. 109–135. Special Issue on Industrial Geometry (Guest editors: B. Jüttler, H. Pottmann, O. Scherzer).

[6] D. Butnariu and Alfredo N. Iusem, *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization*, vol. 40 of Applied Optimization, Kluwer Academic, Dordrecht, 2000.

[7] Emmanuel J. Candes and Terence Tao, *The Dantzig selector: statistical estimation when p is much larger than n*, Ann. Statist., 35 (2007), pp. 2313–2351.

[8] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders, *Atomic decomposition by basis pursuit*, SIAM Journal on Scientific Computing, 20 (1998), pp. 33–61.

[9] Ingrid Daubechies, Michel Defrise, and Christine De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Communications in Pure and Applied Mathematics, 57 (2004), pp. 1413–1457.

[10] Ivar Ekeland and Roger Temam, *Convex analysis and variational problems*, vol. 1 of Studies in Mathematics and its Applications, North-Holland Publishing Co., Amsterdam-Oxford, 1976.

[11] Heinz W. Engl, Martin Hanke, and Andreas Neubauer, *Regularization of inverse problems*, vol. 375 of Mathematics and its Applications, Kluwer Academic Publishers Group, Dordrecht, 1996.

[12] Klaus Frick and Otmar Scherzer, *Regularization of ill-posed linear equations by the non-stationary Augmented Lagrangian Method*, J. Integral Equations Appl., 22 (2010), pp. 217–258.

[13] Markus Grasmair, Markus Haltmeier, and Otmar Scherzer, *Sparse regularization with ℓ^q penalty term*, Inverse Problems, 24 (2008), p. 055020 (13pp).

[14] ——, *Necessary and sufficient conditions for linear convergence of ℓ^1-regularization*, Comm. Pure Appl. Math, (2010).

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

[16] Magnus R. Hestenes, *Multiplier and gradient methods*, J. Optimization Theory Appl., 4 (1969), pp. 303–320.

[17] A.V. Kryanev, *An iterative method for solving incorrectly posed problems*, USSR Computational Mathematics and Mathematical Physics, 14 (1974), pp. 24–33.

[18] Dirk A. Lorenz, *Convergence rates and source conditions for Tikhonov regularization with sparsity constraints*, Journal of Inverse and Ill-Posed Problems, 16 (2008), pp. 463–478.

[19] Dirk A. Lorenz, Stefan Schiffler, and Dennis Trede, *Beyond convergence rates: Exact inversion with Tikhonov regularization with sparsity constraints*. Submitted for publication, 2010.

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

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

[22] R. Ramlau and E. Resmerita, *Convergence rates for regularization with sparsity constraints*, ETNA, 37 (2010), pp. 87–104.

[23] R. Tyrrell Rockafellar, *Augmented Lagrange multiplier functions and duality in nonconvex programming*, SIAM J. Control, 12 (1974), pp. 268–285. Collection of articles dedicated to the memory of Lucien W. Neustadt.

[24] Leonid I. Rudin, Stanley J. Osher, and Emad Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.