ÖAW AUSTRIAN ACADEMY OF SCIENCES

Johann Radon Institute for
Computational and Applied Mathematics
Austrian Academy of Sciences (ÖAW)

RICAM JOHANN-RADON-INSTITUTE FOR COMPUTATIONAL AND APPLIED MATHEMATICS

Convergence rates for Morozov's Discrepancy Principle using Variational Inequalities

S. Anzengruber, R. Ramlau

RICAM-Report 2011-06

Convergence rates for Morozov's Discrepancy Principle using Variational Inequalities

Stephan W Anzengruber* and Ronny Ramlau*†

Abstract

We derive convergence rates for Tikhonov-type regularization with convex penalty terms, where the regularization parameter is chosen according to Morozov's discrepancy principle and variational inequalities are used to generalize classical source and nonlinearity conditions. Rates are obtained first with respect to the Bregman distance and a Taylor-type distance and those results are combined to derive rates in norm and the penalty term topology.

For the special case of the sparsity promoting weighted ℓ_p-norms as penalty terms and for a searched-for solution, which is known to be sparse, the above results give convergence rates of up to linear order.

Keywords. Inverse problems, Morozov's discrepancy principle, Variational inequalities, Convergence rates, Regularization, Sparsity.

AMS subject classification. 47J06; 65J20; 49N45.

1 Introduction

Many problems arising in phyiscal applications can be modeled mathematically as an operator equation

$$F(x) = y, \tag{1}$$

where one is interested in finding a quantity x from observed data y. Examples include, but are by no means limited to, medical and astronomical imaging, inverse scattering and mathematical finance. Frequently the data y will be corrupted by noise, for instance, if they were obtained through a measurement process which is subject to inaccuracy. We will indicate the noisy version of the data by y^δ. If the operator under consideration is ill-posed, even small data errors may lead to large errors in the reconstruction.

If the data y, y^δ belong to a normed space Y, as a first step towards making problem (1) mathematically more tangible, one can consider instead the minimization of the least-squares functional

$$J(x) = \|F(x) - y^\delta\|_Y^2, \tag{2}$$

*Johann Radon Institute for Computational and Applied Mathematics, Altenberger Strasse 69, A-4040 Linz, Austria. Email: stephan.anzengruber@ricam.oeaw.ac.at

†Johannes Kepler University Linz, Institute for Industrial Mathematics, Altenberger Strasse 69, A-4040 Linz, Austria

and denote the noise level by δ, i.e., $\|y - y^\delta\| \leq \delta$. This formulation allows for the definition of a generalized solution even if the noisy data do not belong to the range of the operator, $y^\delta \notin \operatorname{rg} F$, but problem (2) remains ill-posed. One approach to overcome the aforementioned difficulties of ill-posed problems is to replace (2) by a family $\{J_\alpha\}_{\alpha>0}$ of neighboring well-posed (or at least stable) problems, which incorporate additional a-priori knowledge of properties of the searched-for solution $x^\dagger$ through a regularizing functional $\Psi(x)$. For the purpose of this paper we will only be concerned with convex Ψ. The approximate solutions are taken to be the minimizers – denoted by x_α^δ – of the resulting variational functional

$$J_\alpha(x) = \|F(x) - y^\delta\|_Y^2 + \alpha\Psi(x). \tag{3}$$

The choice $\Psi(x) = \|x\|_X^2$, for x belonging to some Hilbert space X, constitutes the classical Tikhonov regularization. We refer the reader to [5] for further details in this respect.

Other choices which have received considerable attention in recent years, are total variation and sparsity promoting weighted ℓ_p-norms with respect to a given basis or frame $\{\phi_\lambda\}_{\lambda\in\Lambda} \subset X$,

$$\Psi_{w,p}(x) = \sum_{\lambda\in\Lambda} w_\lambda \, |\langle \phi_\lambda, x\rangle|^p, \qquad 0 < w_0 \leq w_\lambda, 1 \leq p \leq 2. \tag{4}$$

Sparse representations of solutions are of strong interest, for example, in signal compression and astronomical imaging, where objects of interest like images are sparse, but their standard reconstructions are not. Enforcing sparsity adds knowledge on the solution and therefore improves the reconstruction.

The choice of the regularization parameter α in (3) turns out to be of crucial importance for the quality of the resulting reconstructions. Many strategies have been proposed in the literature and they typically lead to a somewhat different behaviour in terms of convergence and, especially, rates of convergence in the chosen topology of the regularized solutions $x_\alpha^\delta \to x^\dagger$ as $\delta \to 0$, where $x^\dagger$ denotes the searched-for solution to problem (1). For *a-priori* parameter choice rules, where $\alpha = \alpha(\delta)$ depends on the noise level δ only, convergence rates have been shown with respect to the Bregman distance in [4, 18, 19], and for the functionals $\Psi_{w,p}$ in (4) these results could even be used to obtain convergence rates in norm (cf. [7, 13, 17]).

When using Morozov's Discrepancy Principle (henceforth, MDP), which belongs to the class of *a-posteriori* parameter choice rules, i.e., the regularization parameter depends not only on the noise level, but also on the noisy data y^δ, we choose $\alpha = \alpha(\delta, y^\delta)$ such that for some minimizer x_α^δ of (3)

$$\tau_1\delta \leq \left\| F\left(x^\delta_{\alpha(\delta,y^\delta)}\right) - y^\delta \right\| \leq \tau_2\delta, \qquad 1 \leq \tau_1 \leq \tau_2 \tag{5}$$

holds. This way of choosing the regularization parameter has been studied in great detail in its present formulation [1, 3, 15] as well as in several related variations [5, 8, 12, 14, 21].

It has been shown in [5] that for linear operator equations and certain classes of regularization methods defined via spectral decomposition in Hilbert spaces the discrepancy principle gives order optimal convergence rates. These results cover the classical Tikhonov regularization mentioned above, but not variational regularization methods with general convex penalty terms as in (3), and in particular not the functionals $\Psi_{p,w}$ in (4) for $p < 2$.

For the special case of denoising, where the operator under consideration is the identity in $L^2(\mathbb{R}^d)$, with L^1 or ℓ_1-penalty term, optimal order convergence rate results were obtained in [12]. It was also shown that the resulting regularization method does not saturate, in which case the discrepancy principle yields the same convergence rates as a-priori parameter choice rules.

For Tikhonov-type regularization of linear operator equations with general convex penalty terms as in (3) and regularization parameter chosen according to MDP, Bonesky [3] showed convergence rate results with respect to the Bregman distance in reflexive Banach spaces and his results were generalized to non-linear operators in [1] adding an additional condition on the structure of the non-linearity in F.

Finally, the residual method was studied in the report [8]. It is closely related to the discrepancy priniple and in r-convex Banach spaces, $r \geq 2$, convergence rates in norm were derived when using the penalty term $\Psi(x) = \|x\|_X^r/r$. Moreover, for linear operators additional convergence rates were provided under the assumption that the unknown solution is sparse.

In the present paper we study non-linear operators in Banach spaces and obtain convergence rates in norm of up to linear order with respect to the data error using source and non-linearity conditions formulated through variational inequalities in combination with Morozov's discrepancy principle. We show that if the searched-for solution $x^\dagger$ is sparse, then a linear convergence rate can be obtained when penalizing with $\Psi_{w,1}$ as defined in (4).

The paper is structured as follows. In Section 2 we specify our setting. Then we derive the main results about convergence rates both with respect to the Bregman distance and in norm in Section 3. Sparse recovery in Hilbert spaces is studied in Section 4 as an example, which will be found to be a special case of the framework described in Section 3, and we show that convergence rates of up to linear order can be observed for non-linear operators using MDP in combination with variational inequalities. Finally, Section 5 provides a discussion which illustrates the link between the variational inequalities used in our analysis and classical source and non-linearity conditions.

2 Preliminaries

Throughout this paper we assume the operator $F : \mathrm{dom}(F) \subset X \to Y$, with $0 \in \mathrm{dom}(F)$, to be weakly continuous between reflexive Banach spaces X and Y with dual spaces X^* and Y^*, respectively, and that the penalty term $\Psi(x)$ fulfills the following

Condition 2.1. Let $\Psi : \mathrm{dom}(\Psi) \subset X \to \mathbb{R}^+$, with $0 \in \mathrm{dom}(\Psi)$, be a convex functional such that

(i) $\Psi(x) = 0$ if and only if $x = 0$,

(ii) Ψ is weakly lower semicontinous w.r.t. the norm in X,

(iii) Ψ is weakly coercive, i.e. $\|x_n\| \to \infty \implies \Psi(x_n) \to \infty$.

We will frequently encounter the following example of penalty terms, which fulfill the above condition, throughout this paper.

Example 2.2. Let $\Phi = \{\phi_\lambda\}_{\lambda\in\Lambda}$ be a Schauder basis for X and denote by $\mathbf{x} = \{x_\lambda\}_{\lambda\in\Lambda}$ the coefficients of $x \in X$ w.r.t. Φ. Then, for a fixed sequence $w = \{w_\lambda\}_{\lambda\in\Lambda}$ which is assumed to be bounded from below away from zero, $0 < w_0 \le w_\lambda$ for all $\lambda \in \Lambda$ and for $1 \le p \le 2$, we define

$$\Psi_{p,w}(x) = \sum_{\lambda\in\Lambda} w_\lambda \, |x_\lambda|^p \,, \tag{6}$$

with

$$\mathrm{dom}(\Psi_{p,w}) = \{x \in X : \Psi_{p,w}(x) < \infty\}.$$

At this point we would like to fix some notational conventions.

Definition 2.3. We denote the set of all Ψ-minimizing solutions of $F(x) = y$ by $\mathcal{L}$, i.e.

$$\mathcal{L} = \{x^\dagger \in X |\ F\left(x^\dagger\right) = y \text{ and } \Psi(x^\dagger) \le \Psi(x)\ \forall x \text{ s.t. } F(x) = y\}, \tag{7}$$

and we assume $\mathcal{L} \ne \emptyset$.

Our regularization method consists in minimizing Tikhonov-type variational functionals defined as

$$J_\alpha(x) = \begin{cases} \|F(x) - y^\delta\|^q + \alpha\Psi(x) & \text{if} \quad x \in \mathcal{D} \\ +\infty & \text{otherwise,} \end{cases} \tag{8}$$

where $\mathcal{D} := \mathrm{dom}(F) \cap \mathrm{dom}(\Psi)$ and $q > 0$ is fixed. Hence, the regularized solutions are chosen to be minimizers of these functionals,

$$x_\alpha^\delta \in \mathcal{M}_\alpha = \arg\min_{x\in X}\{J_\alpha(x)\}. \tag{9}$$

In general, the minimizers of (8) will not be unique.

Now, let us come to the parameter choice rule of interest to us.

Definition 2.4. When using Morozov's Discrepancy Principle (MDP) we choose the regularization parameter $\alpha = \alpha(\delta, y^\delta)$ such that for constants $1 < \tau_1 \le \tau_2$

$$\tau_1\delta \le \|F\left(x_\alpha^\delta\right) - y^\delta\| \le \tau_2\delta. \tag{10}$$

It has been shown in [1] that the following condition is sufficient for the existence of $\alpha = \alpha(\delta, y^\delta)$ and $x_\alpha^\delta \in \mathcal{M}_\alpha$ fulfilling (10).

Condition 2.5. Assume that y^δ satisfies

$$\|y - y^\delta\| \le \delta < \tau_2\delta < \|F(0) - y^\delta\|, \tag{11}$$

and that there is no $\alpha > 0$ with minimizers $x_1, x_2 \in M_\alpha$ such that

$$\|F(x_1) - y^\delta\| < \tau_1\delta \le \tau_2\delta < \|F(x_2) - y^\delta\|.$$

For the purpose of this paper we will assume, henceforth, that α, x_α^δ as in (10) can indeed be found, which is certainly the case if Condition 2.5 holds true.

Remark 2.6. An immediate consequence of (10) which we will need repeatedly, is that

$$\|F\left(x_\alpha^\delta\right) - F\left(x^\dagger\right)\| \le \|F\left(x_\alpha^\delta\right) - y^\delta\| + \|y^\delta - y\| \le (\tau_2 + 1)\delta. \tag{12}$$

The next Lemma can be found in [1], we give a proof here for the convenience of the reader.

Lemma 2.7. *If α is chosen according to MDP, then*

$$\Psi(x_\alpha^\delta) \leq \Psi(x^\dagger)$$

holds for all $x^\dagger \in \mathcal{L}$ and $x_\alpha^\delta \in \mathcal{M}_\alpha$ satisfying (10).

Proof. Using (10) and the minimizing property of $x_\alpha^\delta \in \mathcal{M}_\alpha$ we see that

$$\tau_1^q \delta^q + \alpha\Psi(x_\alpha^\delta) \leq \|F\left(x_\alpha^\delta\right) - y^\delta\|^q + \alpha\Psi(x_\alpha^\delta) \leq \delta^q + \alpha\Psi(x^\dagger).$$

For $\tau_1 \geq 1$ we thus get

$$0 \leq (\tau_1^q - 1)\frac{\delta^q}{\alpha} \leq \Psi(x^\dagger) - \Psi(x_\alpha^\delta),$$

which completes the proof.

□

3 Convergence rates

In order to formulate the variational inequalities as well as to measure and estimate convergence rates, we will make use of the Bregman distance.

Definition 3.1. Let $\partial\Psi(x) \subset X^*$ denote the subgradient of Ψ at $x \in X$. The generalized Bregman distance with respect to Ψ of two elements $x, z \in X$ is defined as

$$D_\Psi(x, z) = \{D_\Psi^\xi(x, z) : \xi \in \partial\Psi(z) \neq \emptyset\},$$

where

$$D_\Psi^\xi(x, z) = \Psi(x) - \Psi(z) - \langle \xi, x - z\rangle \tag{13}$$

denotes the Bregman distance with respect to Ψ and $\xi \in \partial\Psi(z)$. We remark that from here on $\langle\cdot,\cdot\rangle$ denotes the dual pairing in X^*, X or Y^*, Y and not the inner product on a Hilbert space (unless noted otherwise).

Throughout this paper we will assume that the operator $F : X \to Y$ is Gâteaux differentiable at arbitrary but fixed $x^\dagger \in \mathcal{L}$ and that the derivative $F'(x^\dagger) : X \to Y$ is bounded. We start by introducing the following notational conventions.

Definition 3.2. For $x \in X, x^\dagger \in \mathcal{L}$, we denote the norm of the second order Taylor remainder by

$$\mathcal{T}(x, x^\dagger) = \|F(x) - F\left(x^\dagger\right) - F'(x^\dagger)(x - x^\dagger)\|, \tag{14}$$

and call

$$D_\mathcal{T}(x, x^\dagger) = \|F'(x^\dagger)(x - x^\dagger)\| \tag{15}$$

the *Taylor distance* of x and $x^\dagger$.

In the recent work [2] Boţ and Hofmann formulated the conjecture, that convergence rate results cannot be proven if the operator F fails to satisfy a structural condition of the form

$$D_{\mathcal{T}}(x, x^\dagger) \le C\sigma(\, \|F(x) - F\left(x^\dagger\right)\|), \tag{16}$$

where σ is a continuous, strictly increasing function through the origin. We now introduce variational inequalities which are generalizations of the standard source and nonlinearity conditions discussed in Example 5.1 below, and which ultimately also fall into the framework of (16). Similar inequalities were also used in the recent works [2, 6, 7, 11].

Condition 3.3. (Variational inequalities) For $\xi \in \partial\Psi(x^\dagger)$ and $0 < \kappa \le 1$, assume that there are $\rho > 0$ and $\beta_i, \gamma_i \ge 0, i = 1, 2, 3,$ such that for all $x \in \mathcal{B}_\rho(x^\dagger) \cap \mathcal{D}$ it holds that

$$-\langle \xi, x - x^\dagger \rangle \le \beta_1 D_\Psi^\xi(x, x^\dagger) + \beta_2 D_{\mathcal{T}}(x, x^\dagger) + \beta_3 \|F(x) - F\left(x^\dagger\right)\|^\kappa \tag{17}$$

$$\mathcal{T}(x, x^\dagger) \le \gamma_1 D_\Psi^\xi(x, x^\dagger) + \gamma_2 D_{\mathcal{T}}(x, x^\dagger) + \gamma_3 \|F(x) - F\left(x^\dagger\right)\|^\kappa, \tag{18}$$

where the constants β_i, γ_i fulfill

$$\beta_1 < 1, \qquad \gamma_2 < 1 \quad \text{and} \quad \frac{\beta_2\gamma_1}{(1-\beta_1)(1-\gamma_2)} \le 1. \tag{19}$$

Examples of applications where variational inequalities as in Condition 3.3 are fulfilled are phase retrieval problems and inverse option pricing, which were studied in [9].

Remark 3.4. (i) If the operator F under consideration is linear, (18) becomes a tautology and

$$D_{\mathcal{T}}(x, x^\dagger) = \|F(x) - F\left(x^\dagger\right)\|.$$

Therefore, this term does not contribute to the estimates as it is dominated by $\|F(x) - F\left(x^\dagger\right)\|^\kappa$.

(ii) To prove convergence rates results, the inequalities in Condition 3.3 will only ever need to be evaluated at points $x = x_\alpha^\delta$. Therefore, when using MDP as the parameter choice rule, under assumption (19) it suffices to consider (17) and (18) for the special case $\beta_1 = \gamma_2 = 0$:

Indeed, if (17) holds, then using Lemma 2.7 we obtain

$$\begin{aligned}
&-\langle \xi, x_\alpha^\delta - x^\dagger \rangle \\
&\quad \le \beta_1'(\Psi(x_\alpha^\delta) - \Psi(x^\dagger)) + \beta_2' D_{\mathcal{T}}(x_\alpha^\delta, x^\dagger) + \beta_3' \|F\left(x_\alpha^\delta\right) - F\left(x^\dagger\right)\|^\kappa \\
&\quad \le \beta_2' D_{\mathcal{T}}(x_\alpha^\delta, x^\dagger) + \beta_3' \|F\left(x_\alpha^\delta\right) - F\left(x^\dagger\right)\|^\kappa,
\end{aligned}$$

where

$$\beta_i' = \frac{\beta_i}{1-\beta_1}, \quad \text{for} \quad i = 1, 2, 3.$$

Similarly, if (18) holds, then one gets

$$\begin{aligned}
&\mathcal{T}(x_\alpha^\delta, x^\dagger) \\
&\quad \le \gamma_1' D_\Psi^\xi(x_\alpha^\delta, x^\dagger) + \gamma_2' \|F\left(x_\alpha^\delta\right) - F\left(x^\dagger\right)\| + \gamma_3' \|F\left(x_\alpha^\delta\right) - F\left(x^\dagger\right)\|^\kappa \\
&\quad \le \gamma_1' D_\Psi^\xi(x_\alpha^\delta, x^\dagger) + (\gamma_2' + \gamma_3') \|F\left(x_\alpha^\delta\right) - F\left(x^\dagger\right)\|^\kappa,
\end{aligned}$$

where the last estimation holds whenever $\delta < 1/\tau_2$ and thus asymptotically as $\delta \to 0$, which is the case of interest.

In [1] it has been proven that for the parameter choice rule MDP, the source condition from Example 5.1 (i) and a nonlinearity condition as in (35) yield a convergence rate of order $\mathcal{O}(\delta)$ in the Bregman distance. We will now show that similar results still hold under the more general Condition 3.3 with respect to the Bregman distance and also in the Taylor distance $D_{\mathcal{T}}(x, x^\dagger)$.

Theorem 3.5. *Let Condition 3.3 hold for $x^\dagger \in \mathcal{L}, \xi \in \partial\Psi(x^\dagger)$. If $\alpha = \alpha(\delta, y^\delta)$ is chosen according to MDP then for $x_\alpha^\delta \in \mathcal{M}_\alpha$ satisfying (10), it holds that*

$$D_\Psi^\xi(x_\alpha^\delta, x^\dagger) = \mathcal{O}(\delta^\kappa) \quad as \quad \delta \to 0, \tag{20}$$

$$D_{\mathcal{T}}(x_\alpha^\delta, x^\dagger) = \mathcal{O}(\delta^\kappa) \quad as \quad \delta \to 0. \tag{21}$$

Proof. Following the reasoning in Remark 3.4 (ii), in (17) and (18) we only need to consider the special case where $\beta_1 = \gamma_2 = 0$. Combining the two inequalities and using Lemma 2.7, we then find

$$\begin{aligned} &D_\Psi^\xi(x_\alpha^\delta, x^\dagger) = \Psi(x_\alpha^\delta) - \Psi(x^\dagger) - \langle \xi, x_\alpha^\delta - x^\dagger \rangle \\ &\leq \Psi(x_\alpha^\delta) - \Psi(x^\dagger) + \beta_2 D_{\mathcal{T}}(x_\alpha^\delta, x^\dagger) + \beta_3 \left\| F\left(x_\alpha^\delta\right) - F\left(x^\dagger\right) \right\|^\kappa \\ &\leq \beta_2 \left\| F\left(x_\alpha^\delta\right) - F\left(x^\dagger\right) \right\| + \beta_2\gamma_1 D_\Psi^\xi(x_\alpha^\delta, x^\dagger) + (\beta_2\gamma_3 + \beta_3) \left\| F\left(x_\alpha^\delta\right) - F\left(x^\dagger\right) \right\|^\kappa. \end{aligned}$$

Thus, (19) allows to employ a bootstrap argument, which together with the definition of MDP in (10) gives

$$\begin{aligned} D_\Psi^\xi(x_\alpha^\delta, x^\dagger) &\leq \frac{\beta_2}{1 - \beta_2\gamma_1} \left\| F\left(x_\alpha^\delta\right) - F\left(x^\dagger\right) \right\| + \frac{\beta_2\gamma_3 + \beta_3}{1 - \beta_2\gamma_1} \left\| F\left(x_\alpha^\delta\right) - F\left(x^\dagger\right) \right\|^\kappa \\ &\leq \frac{\beta_2}{1 - \beta_2\gamma_1}\tau_2\delta + \frac{\beta_2\gamma_3 + \beta_3}{1 - \beta_2\gamma_1}\tau_2^\kappa\delta^\kappa = \mathcal{O}(\delta^\kappa) \quad \text{as} \quad \delta \to 0. \end{aligned}$$

Knowing this, we obtain from (18) using again (10)

$$\begin{aligned} D_{\mathcal{T}}(x_\alpha^\delta, x^\dagger) &\leq \left\| F\left(x_\alpha^\delta\right) - F\left(x^\dagger\right) \right\| + \gamma_1 D_\Psi^\xi(x_\alpha^\delta, x^\dagger) + \gamma_3 \left\| F\left(x_\alpha^\delta\right) - F\left(x^\dagger\right) \right\|^\kappa \\ &\leq \frac{1}{1 - \beta_2\gamma_1}\tau_2\delta + \frac{\gamma_1\beta_3 + \gamma_3}{1 - \beta_2\gamma_1}\tau_2^\kappa\delta^\kappa = \mathcal{O}(\delta^\kappa) \quad \text{as} \quad \delta \to 0. \end{aligned}$$

□

To prove convergence rates with respect to the topology induced by the penalty term we introduce another variational inequality.

Condition 3.6. Let $x^\dagger \in \mathcal{L}$, $\xi \in \partial\Psi(x^\dagger)$ and assume there exist $\mu_i \geq 0, r, \rho > 0$, and $0 < \kappa \leq 1$ such that for all $x \in B_\rho(x^\dagger) \cap \mathcal{D}$ it holds

$$\Psi(x - x^\dagger)^r \leq \mu_1 D_\Psi^\xi(x, x^\dagger) + \mu_2 D_{\mathcal{T}}(x, x^\dagger) + \mu_3 \left\| F\left(x\right) - F\left(x^\dagger\right) \right\|^\kappa. \tag{22}$$

In Section 4 below, we will see that in the context of sparse recovery Condition 3.6 is satisfied. When using MDP as the parameter choice rule the additional variational inequalitiy (22) immediately yields convergence rates.

Theorem 3.7. *If Conditions 3.3 and 3.6 hold for $x^\dagger \in \mathcal{L}$ and $\alpha = \alpha(\delta, y^\delta)$ is chosen according to MDP, then*

$$\Psi(x_\alpha^\delta - x^\dagger) = \mathcal{O}(\delta^{\kappa/r}) \quad \textit{as} \quad \delta \to 0 \tag{23}$$

holds for any $x_\alpha^\delta \in \mathcal{M}_\alpha$ satisfying (10).

Proof. The assumptions of Theorem 3.5 hold for $x = x_\alpha^\delta$ and from (22), (20), (21), and (10) we get

$$\begin{aligned}\Psi(x_\alpha^\delta - x^\dagger)^r &\le \mu_1 D_\Psi^\xi(x_\alpha^\delta, x^\dagger) + \mu_2 D_\mathcal{T}(x_\alpha^\delta, x^\dagger) + \mu_3 \left\|F\left(x_\alpha^\delta\right) - F\left(x^\dagger\right)\right\|^\kappa \\ &= \mathcal{O}(\delta^\kappa),\end{aligned}$$

which is the desired convergence rate.

□

Remark 3.8. It is worthwhile noting, that if one has

$$\|x - x^\dagger\|_X^r \le \mu_1 D_\Psi^\xi(x, x^\dagger) + \mu_2 D_\mathcal{T}(x, x^\dagger) + \mu_3 \left\|F(x) - F\left(x^\dagger\right)\right\|^\kappa. \tag{24}$$

instead of (22), then in complete analogy to Theorem 3.7 one obtains a convergence rate in norm, namely

$$\|x_\alpha^\delta - x^\dagger\|_X = \mathcal{O}(\delta^{\kappa/r}) \quad \text{as} \quad \delta \to 0. \tag{25}$$

This would be the case, for example, if $\Psi(x)$ is locally q-coercive for $2 \le q < \infty$, i.e., for some $c_q, \rho > 0$,

$$\|x - x^\dagger\|_X^q \le c_q D_\Psi(x, x^\dagger) \tag{26}$$

holds for all $x \in \mathcal{B}_\rho(x^\dagger)$. It is well known, that the sparsity constraints $\Psi_{p,w}(x)$ defined in (6) fulfill (26) with $q = 2$, and for the optimal case $\kappa = 1$ we would obtain the classical rate

$$\|x_\alpha^\delta - x^\dagger\|_X = \mathcal{O}(\delta^{1/2}) \quad \text{as} \quad \delta \to 0. \tag{27}$$

But – as we will see in Section 4 – even (22) holds true in this setting and the resulting convergence result with respect to $\Psi_{p,w}$ is stronger than convergence in norm, which is why we prefer to work with formulation (22). Nevertheless, for different choices of the penalty term Ψ it may be more suitable to use (24) instead.

4 Sparse recovery

As a prominent case study, we will show now that the convergence rate from (27) for the sparsity promoting penalty terms $\Psi_{p,w}$ defined in (6) can be improved significantly when including the a-priori information that the $\Psi_{p,w}$-minimizing solution $x^\dagger$ is also sparse. To this end we show that a variational inequality as in Condition 3.6 holds true for this method.

Condition 4.1. Throughout this section we assume that X is a Hilbert space and that $x^\dagger \in \mathcal{L}$ is sparse with respect to a fixed frame $\{\phi_\lambda\}_{\lambda\in\Lambda}$ for X.

For elements $x \in X$ we use the shorthand notation

$$x_\lambda = \langle \phi_\lambda, x\rangle.$$

Then, for fixed $\xi \in \partial\Psi(x^\dagger) \subset X^* = X$, the set

$$J = \{\lambda \in \Lambda \mid x^\dagger_\lambda \neq 0 \vee |\xi_\lambda| \geq w_0\} \tag{28}$$

is finite, since $x^\dagger$ is sparse and the sequence $\{\xi_\lambda\}_{\lambda\in\Lambda}$ belongs to $\ell_2(\Lambda)$. We denote the subspace spanned by elements with indices in J by

$$U = \operatorname{span}\{\phi_\lambda \mid \lambda \in J\},$$

and the projections of X onto U and $U^\perp$ by π and $\pi^\perp$, respectively.

Moreover, we assume that $F'(x^\dagger)|_U$ is injective, i.e., for all $x, z \in X$ from $F'(x^\dagger)(x-z) = \pi^\perp(x-z) = 0$ if follows that $x = z$. Note that $F'(x^\dagger)|_U$ is clearly injective, if $F'(x^\dagger)$ satisfies the so-called *FBI property* (see, e.g., [13] and the references therein for further information).

A variational inequality of type (22) indeed holds in the sparse recovery case which will allows us to derive convergence rates whenever Condion 3.3 is satisfied. This is shown in Theorem 4.3 and Corollary 4.4 below. The proof is based on techniques from [7] and uses the following technical Lemma.

Lemma 4.2. *If condition 4.1 is satisfied, then there exists $c > 0$ such that for any $x \in X, \xi \in \partial\Psi_{p,w}(x^\dagger)$ and $\lambda \notin J$*

$$w_\lambda\, |x_\lambda|^p \leq c\, (w_\lambda\, |x_\lambda|^p - \xi_\lambda x_\lambda) \tag{29}$$

holds.

Proof. We first consider the case $p > 1$, then the unique element in the subgradient $\xi \in \partial\Psi_{p,w}(x^\dagger)$,

$$\xi = \sum_{\lambda\in\Lambda} p\, w_\lambda \operatorname{sign}(\langle\phi_\lambda, x^\dagger\rangle)\, \left|\langle\phi_\lambda, x^\dagger\rangle\right|^{p-1} \phi_\lambda,$$

satisfies $\xi_\lambda = 0$ whenever $x^\dagger_\lambda = 0$, which in turn holds for all $\lambda \notin J$, so that (29) clearly holds for all $c \geq 1$.

On the other hand, for $p = 1$ we define

$$m = \max_{\lambda\notin J} |\xi_\lambda| < w_0.$$

Here the maximum is attained since $\xi \in X^* = X$ and thus the sequence $\{\xi_\lambda\}_{\lambda\in\Lambda}$ belongs to $\ell_2(\Lambda)$. Since $0 \leq |\xi_\lambda| \leq m < w_0 \leq w_\lambda$ the choice $c = 2w_0/(w_0 - m)$ yields

$$\begin{aligned}\frac{w_\lambda}{c}\, |x_\lambda| &= w_\lambda\, |x_\lambda| - \frac{w_\lambda}{w_0}\,\frac{w_0+m}{2}\, |x_\lambda| \\ &\leq w_\lambda\, |x_\lambda| - m\, |x_\lambda| \leq w_\lambda\, |x_\lambda| - \xi_\lambda x_\lambda,\end{aligned}$$

and (29) follows.

□

We now show that a variational inequality (22) holds in the sparse recovery case.

Theorem 4.3. *If condition 4.1 is satisfied, then for* $\rho < \|F'(x^\dagger)\|^{-1}$ *and* $x \in B_\rho(x^\dagger)$ *it holds*

$$\Psi_{p,w}(x - x^\dagger) \leq \mu_1 D^{\xi}_{\Psi_{p,w}}(x, x^\dagger) + \mu_2 D_{\mathcal{T}}(x, x^\dagger), \qquad 1 \leq p \leq 2. \tag{30}$$

Proof. In order to estimate the difference between x and $x^\dagger$ with respect to the penalty term, we use the splitting

$$\Psi_{p,w}(x - x^\dagger) = \sum_{\lambda \in J} w_\lambda \left| x_\lambda - x^\dagger_\lambda \right|^p + \sum_{\lambda \notin J} w_\lambda \left| x_\lambda - x^\dagger_\lambda \right|^p, \tag{31}$$

and write

$$c_w = \sup_{\lambda \in J} \{w_\lambda\},$$

which is a finite number because the set J, defined in (28), is finite. Using the equivalence of norms on finite dimensional spaces, we find a constant c_p such that

$$\begin{aligned}
\sum_{\lambda \in J} w_\lambda \left| x_\lambda - x^\dagger_\lambda \right|^p &\leq c_w \left\| \{x_\lambda - x^\dagger_\lambda\}_{\lambda \in J} \right\|^p_{\ell_p(J)} \\
&\leq c_w c_p \left\| \{x_\lambda - x^\dagger_\lambda\}_{\lambda \in J} \right\|^p_{\ell_2(J)} \\
&= c_w c_p \left\| \pi(x - x^\dagger) \right\|^p
\end{aligned}$$

Due to the injectivity of $F'(x^\dagger)$ on U, the boundedness of $F'(x^\dagger)$ and the inequality $(a+b)^p \leq 2(a^p + b^p)$ for $a, b \geq 0$ and $p \geq 1$, we get the following estimate.

$$\begin{aligned}
\|\pi(x - x^\dagger)\|^p &\leq c' \left\| F'(x^\dagger)\pi(x - x^\dagger) \right\|^p \\
&\leq 2c' \left(\|F'(x^\dagger)(x - x^\dagger)\|^p + \|F'(x^\dagger)\|^p \, \|\pi^\perp x\|^p \right).
\end{aligned}$$

From the well known inequality $\|.\|_{\ell_2} \leq \|.\|_{\ell_p}$ for $1 \leq p \leq 2$, Lemma 4.2 and $x^\dagger_\lambda = 0$ for all $\lambda \notin J$, it follows that

$$\begin{aligned}
\|\pi^\perp x\|^p = \Big(\sum_{\lambda \notin J} |x_\lambda|^2 \Big)^{p/2} &\leq \sum_{\lambda \notin J} \frac{w_\lambda}{w_0} \, |x_\lambda|^p \\
&= \frac{c}{w_0} \sum_{\lambda \notin J} w_\lambda \, |x_\lambda|^p - w_\lambda \left| x^\dagger_\lambda \right|^p - \xi_\lambda (x_\lambda - x^\dagger_\lambda) \\
&\leq \frac{c}{w_0} D^{\xi}_{\Psi_{p,w}}(x, x^\dagger),
\end{aligned}$$

where the last inequality holds because all remaining summands for $\lambda \in J$ are Bregman distances $D^{\xi_\lambda}_{w_\lambda\,|.|}(x_\lambda, x^\dagger_\lambda)$, where $\xi_\lambda \in \partial(w_\lambda \, |.|)(x^\dagger_\lambda)$, and hence nonnegative.

To obtain the remaining estimates for terms corresponding to $\lambda \notin J$ in (31), we again use Lemma 4.2 and $x_\lambda^\dagger = 0$ for all $\lambda \notin J$.

$$\begin{aligned}\sum_{\lambda \notin J} w_\lambda \left| x_\lambda - x_\lambda^\dagger \right|^p &\leq c \sum_{\lambda \notin J} w_\lambda \left| x_\lambda \right|^p - w_\lambda \left| x_\lambda^\dagger \right|^p - \xi_\lambda (x_\lambda - x_\lambda^\dagger) \\ &\leq c \; D^\xi_{\Psi_{p,w}}(x, x^\dagger).\end{aligned}$$

Finally, collecting the above inequalities and using that $D_\mathcal{T}(x, x^\dagger) \leq 1$ whenever $x \in B_\rho(x^\dagger)$, we find that

$$\begin{aligned}\Psi_{p,w}(x - x^\dagger) &= \sum_{\lambda \in J} w_\lambda \left| x_\lambda - x_\lambda^\dagger \right|^p + \sum_{\lambda \notin J} w_\lambda \left| x_\lambda - x_\lambda^\dagger \right|^p \\ &\leq \mu_1 \; \| F'(x^\dagger)(x - x^\dagger) \|^p + \mu_2 \; D^\xi_{\Psi_{p,w}}(x, x^\dagger) \\ &\leq \mu_1 D_\mathcal{T}(x, x^\dagger) + \mu_2 D^\xi_{\Psi_{p,w}}(x, x^\dagger).\end{aligned}$$

holds for all $x \in B_\rho(x^\dagger)$.

□

As a corollary we obtain the convergence rate result.

Corollary 4.4. *If $x^\dagger \in \mathcal{L}$ satisfies Condition 3.3 and 4.1 and $\alpha = \alpha(\delta, y^\delta)$ is chosen according to MDP, then for $x_\alpha^\delta \in \mathcal{M}_\alpha$ satisfying* (10) *we obtain a convergence rate*

$$\Psi_{p,w}(x_\alpha^\delta - x^\dagger) = \mathcal{O}(\delta^\kappa) \quad \textit{as} \quad \delta \to 0. \tag{32}$$

Proof. According to Theorem 4.3 we find that Condition 3.6 is satisfied with $r = 1$ and thus Theorem 3.7 is applicable and provides the result.

□

If we take X to be the sequence space ℓ_2 with the canonical basis, then the penalty terms $\Psi_{p,w}$ are powers of the weighted ℓ_p-norms, namely

$$\Psi_{p,w}(x) = \|x\|_{p,w}^p$$

and therefore (32) corresponds to a convergence rate

$$\|x_\alpha^\delta - x^\dagger\|_{p,w} = \mathcal{O}(\delta^{\kappa/p})$$

and if $\kappa = 1$ we obtain linear convergence speed for ℓ_1-regularization (compare [7, 8]).

5 Relation of variational inequalities to different types of source and nonlinearity conditions

Variational formulations of source and nonlinearity conditions have been used earlier in order to obtain convergence rate results. In this Section, we would like to draw a connection between inequalities (17) and (18) in Condition 3.3 and classical source and nonlinearity conditions. Variational inequalities can be seen as a generalization of the latter as the following examples illustrate.

Example 5.1. (i) If $\xi \in \partial\Psi(x^\dagger)$ fulfills the classical source condition

$$\xi = F'(x^\dagger)^* w, \tag{33}$$

with $w \in Y^*$, then it follows that

$$-\langle \xi, x - x^\dagger \rangle \leq \left| \langle w, F'(x^\dagger)(x - x^\dagger) \rangle \right| \leq \|w\|_{Y*} D_{\mathcal{T}}(x, x^\dagger), \tag{34}$$

and thus (17) holds with $\beta_2 = \|w\|_{Y*}$, and $\beta_1 = \beta_3 = 0$. Note, that the presence of the term $D_{\mathcal{T}}(x, x^\dagger)$ in (17) allows us to express this classical source condition through only the first variational inequality. Omitting this term and using an alternative formulation

$$-\langle \xi, x - x^\dagger \rangle \leq \beta_1 D_\Psi^\xi(x, x^\dagger) + \beta_3 \|F(x) - F\left(x^\dagger\right)\|^\kappa,$$

which has been considered, e.g., in [2, 20], one always needs to also employ some sort of structural nonlinearity condition to include this standard case in the setting.

(ii) One of the first structural assumptions regarding nonlinearity (see, e.g., [5, 16]), was that F be Fréchet differentiable between Hilbert spaces X, Y and that the derivative be locally Lipschitz continuous near $x^\dagger \in \mathcal{L}$, i.e., for some $\rho > 0$ and all $x, z \in \mathcal{B}_\rho(x^\dagger)$ it holds

$$\|F'(x) - F'(z)\| \leq c\, \|x - z\|.$$

Under this assumption one can show that for classical Tikhonov regularization, where $\Psi(x) = \|x\|^2$, locally the following estimate holds

$$\mathcal{T}(x, x^\dagger) \leq \frac{c}{2} \|x - x^\dagger\|^2 = \frac{c}{2} D_\Psi^\xi(x, x^\dagger), \tag{35}$$

which is (18) with $\gamma_1 = c/2$ and $\gamma_2 = \gamma_3 = 0$.

(iii) In [10] an operator F is defined to be *nonlinear of degree* (n_1, n_2, n_3) locally near $x^\dagger$, with $n_1, n_2 \in [0, 1], n_3 \in [0, 2]$, if for some $c, \rho > 0$ and all $x \in B_\rho(x^\dagger)$:

$$\mathcal{T}(x, x^\dagger) \leq c D_{\mathcal{T}}(x, x^\dagger)^{n_1} \|F(x) - F\left(x^\dagger\right)\|^{n_2} \|x - x^\dagger\|^{n_3}.$$

Taking into account the problem under consideration in [10], where X,Y are Hilbert spaces and $\Psi = \|.\|^2$, this definition may be generalized within our framework to

$$\mathcal{T}(x, x^\dagger) \leq c D_{\mathcal{T}}(x, x^\dagger)^{n_1} \|F(x) - F\left(x^\dagger\right)\|^{n_2} D_\Psi^\xi(x, x^\dagger)^{n_3}.$$

Applying Young's inequality twice to that last inequality we find that – whenever $n_1 + n_3 < 1$ – there exist constants γ_i such that (18) holds for x sufficiently close to $x^\dagger$ with

$$\kappa = \min\left(1, \frac{n_2}{1 - n_1 - n_3}\right).$$

From these examples it becomes clear that inequalities (17) and (18) are relaxed versions of the stricter estimates (34) and (35), as they allow for a trade-of between the terms $\|F(x) - F(x^\dagger)\|^\kappa, D_\mathcal{T}(x, x^\dagger)$ and $D_\Psi(x, x^\dagger)$. Note that if one considers examples (i) and (ii) together for $\beta_1 = \gamma_2 = 0$ (cf. Remark 3.4 (ii)), then (19) becomes the well-known smallness condition

$$\frac{c}{2}\|w\|_{Y*} < 1.$$

Regarding the third variational inequality (22) in Condition 3.6, it can be seen as a relaxation of Assumption 1 in [7], where the following formulation was considered: For all x satisfying $\Psi(x) < \rho_1$ and $\|F(x) - F(x^\dagger)\| < \rho_2$, let

$$\Psi(x) - \Psi(x^\dagger) \geq c_1 \|x - x^\dagger\|^r - c_2 \|F(x) - F(x^\dagger)\|, \tag{36}$$

where $\rho_1 > \Psi(x^\dagger)$ and $c_1, c_2, r, \rho_2 > 0$.

One important difference is that in our setting it is sufficient for Condition 3.6 to hold in an arbitrarily small neigbourhood around $x^\dagger$, which is less restrictive than (36).

Moreover, if α is chosen according to MDP, then even locally (36) would be a stronger assumption than (22) when evaluated at points $x = x_\alpha^\delta$, because according to Lemma 2.7

$$\Psi(x_\alpha^\delta) - \Psi(x^\dagger) \leq 0 \leq D_\Psi^\xi(x_\alpha^\delta, x^\dagger) \tag{37}$$

holds in this case.

In [7] also the variational inequalities in Condition 3.3 are formulated with $D_\Psi^\xi(x_\alpha^\delta, x^\dagger)$ replaced by $\Psi(x_\alpha^\delta) - \Psi(x^\dagger)$ and at this point we would like to discuss briefly how this approach can be fitted within the framework of Theorems 3.5 and 3.7. Indeed, for the convergence rate results we only need information at points $x = x_\alpha^\delta$ and from (37) it becomes clear that if such alternative inequalities were to hold instead of the variational inequalities (17), (18) or (22), all estimates in the proofs of Theorems 3.5 and 3.7 are valid in the same way as in the cases $\beta_1 = 0, \gamma_1 = 0$ or $\mu_1 = 0$, respectively, and we obtain the same convergence rates.

Furthermore, if we consider the special case $\gamma_1 = \mu_1 = 0$, then assumption (17) can be dropped entirely as it is only needed to estimate the Bregman distance which no longer appears in our estimates.

If, on the other hand, a stronger nonlinearity condition

$$\mathcal{T}(x, x^\dagger) \leq \gamma_1\big(\Psi(x) - \Psi(x^\dagger)\big) + \gamma_2 D_\mathcal{T}(x, x^\dagger) + \gamma_3 \|F(x) - F(x^\dagger)\| \tag{38}$$

is satisfied locally for $x \in B_\rho(x^\dagger) \subset \mathcal{D}$ (assuming $\mathcal{D}$ contains such a ball) and the penalty term Ψ under consideration is differentiable, then (17) always holds true. As argued in [7] this can be seen by fixing $z \neq 0$ and applying (38) to $z_t = x^\dagger + tz$ (which belongs to $B_\rho(x^\dagger)$ for $t > 0$ small enough) and dividing by t, which yields

$$\begin{aligned}
&\frac{1}{t}\|F(x^\dagger + tz) - F(x^\dagger) - F'(x^\dagger)(tz)\| \\
&\quad \leq \gamma_1 \frac{\Psi(x^\dagger + tz) - \Psi(x^\dagger)}{t} + \gamma_2 \|F'(x^\dagger)z\| + \gamma_3 \frac{1}{t}\|F(x^\dagger + tz) - F(x^\dagger)\|.
\end{aligned}$$

Taking the limit $t \to 0^+$ and choosing $z = x - x^\dagger$ for $x \in X\backslash\{x^\dagger\}$ (note, that if $x = x^\dagger$, then (17) is satisfied trivially) we obtain

$$0 \leq \gamma_1 \langle \Psi'(x^\dagger), x - x^\dagger \rangle + (\gamma_2 + \gamma_3) \|F'(x^\dagger)(x - x^\dagger)\|,$$

which is a special case of (17). This is to say that assumption (38) is strong enough (locally) to ensure that for differentiable penalty terms a variational source condition (17) holds (globally) as well.

Conclusion

We have studied a regularization method for ill-posed, possibly non-linear operator equations through the minimization of a Tikhonov-type functional with general convex penalty term, where the regularization parameter is chosen according to Morozov's discrepancy principle.

If the searched-for solution $x^\dagger$ satisfies a generalized source condition and the operator under consideration a generalized non-linearity condition, which were formulated as variational inequalities, then we found that the difference between the regularized solution obtained through our method and $x^\dagger$ when measured in the Bregman distance or a Taylor-type distance, goes to zero at a rate of δ^κ as $\delta \to 0$, where the parameter $\kappa \in (0, 1]$ allows for a relaxation of the classical source and non-linearity conditions, which are related to the case $\kappa = 1$.

Using another variational inequality (compare (22)), which links the aforementioned Bregman- and Taylor distances to the Banach space norm, we could use the rates established for these distances to obtain a convergence rate $\mathcal{O}(\delta^{\kappa/r})$ in norm. Here the parameter $r \geq \kappa$ stems from the third variational inequality and even though such a constant r may be found from properties of the underlying (Banach) space and the penalty functional alone, it may be improved by additional knowledge about the true solution $x^\dagger$.

This behaviour could be observed when analyzing the situation of a solution which is known to be sparse in a Hilbert space setting, where the penalty term was chosen to be $\Psi_{p,w}$ with $1 \leq p \leq 2$ as defined in (6). A rate with $r = 2$ can always be achieved for these choices, but using the sparsity assumption the third variational inequality could be shown to hold even for $r = p$, which yields convergence rates of up to linear order, $\mathcal{O}(\delta)$, in the limiting case $\kappa = p = 1$.

Acknowledgment

The authors acknowledge support from the Austrian Science Fund (FWF) within the Austrian Academy of Sciences, project P19496-N18.

References

[1] Anzengruber S W and Ramlau R 2010 Morozov's discrepancy principle for Tikhonov-type functionals with nonlinear operators *Inverse Problems* 26(2) 025001 doi:10.1088/0266-5611

[2] Boţ R I and Hofmann B 2010 An extension of the variational inequality approach for obtaining convergence rates in regularization of nonlinear ill-posed problems *Journal of Integral Equations and Applications* 22(3)

[3] Bonesky T 2009 Morozov's discrepancy principle and Tikhonov-type functionals *Inverse Problems* 25(1) 015015 doi:10.1088/0266-5611

[4] Burger M and Osher S 2004 Convergence rates of convex variational regularization *Inverse Problems* 20(5) 1411–21 doi:10.1088/0266-5611/20/5/005

[5] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* vol. 375 of *Mathematics and its Application* (Dordrecht: Kluwer Academic Publishers)

[6] Flemming J and Hofmann B 2010 A new approach to source conditions in regularization with general residual term *Numer. Funct. Anal. Optimiz.* 31 254–284

[7] Grasmair M, Haltmeier M and Scherzer O 2008 Sparse regularization with ℓ^q penalty term *Inverse Problems* 24(5) 1–13

[8] Grasmair M, Haltmeier M and Scherzer O 2009 The Residual Method for Regularizing Ill-posed Problems *Photoacoustic Imaging in Medicine and Biology Report* 14

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

[10] Hofmann B and Scherzer O 1994 Factors influencing the ill-posedness of nonlinear problems *Inverse Problems* 10(6) 1277–97 doi:10.1088/0266-5611/10/6/007

[11] Hofmann B and Yamamoto M 2010 On the interplay of source conditions and variational inequalities for nonlinear ill-posed problems *Applicable Analysis* 89(11) 1705–1727 doi:10.1080/00036810903208148

[12] Justen L and Ramlau R 2009 A general framework for soft-shrinkage with applications to blind deconvolution and wavelet denoising *Appl. and Comp. Harmonic Anal.* 26(1) 43–63

[13] Lorenz D A 2008 Convergence rates and source conditions for Tikhonov regularization with sparsity constraints *Journal of Inverse and Ill-Posed Problems* 16 463–478

[14] Morozov V A 1984 *Methods for solving incorrectly posed problems* (New York: Springer-Verlag)

[15] Ramlau R 2002 Morozov's discrepancy principle for Tikhonov regularization of non-linear operators *Journal for Numer. Funct. Anal. and Opt.* 23 147–172

[16] Ramlau R 2008 Regularization properties of Tikhonov regularization with sparsity constraints *Electron. Trans. Numer. Anal.* 30 54–74

[17] Ramlau R and Resmerita E 2010 Convergence rates for regularization with sparsity constraints *Electronic Transactions on Numerical Analysis* 37 87–104

[18] Resmerita E 2005 Regularization of ill-posed problems in Banach spaces: convergence rates *Inverse Problems* 21 1303–1314

[19] Resmerita E and Scherzer O 2006 Error estimates for non-quadratic regularization and the relation to enhancement *Inverse Problems* 22 801–814

[20] Scherzer O, Grasmair M, Grossauer H, Haltmeier M and Lenzen F 2009 *Variational Methods in Imaging* (New York: Springer-Verlag)

[21] Tikhonov A N, Leonov A S and Yagola A G 1998 *Nonlinear Ill-posed Problems* (London: Chapman & Hall)