Theory and examples of variational regularization with non-metric fitting functionals

Jens Flemming*

July 19, 2010

We describe and analyze a general framework for solving ill-posed operator equations by minimizing Tikhonov-like functionals. The fitting functional may be non-metric and the operator is allowed to be nonlinear and nonsmooth. In comparision to former results on variational regularization with non-metric fitting functionals we significantly weaken the assumptions for proving convergence rates and, in addition, we extend the results to a wider range of rates. Two examples, coming from imaging applications, show that the developed theory is applicable to practically relevant problems.

1 Introduction

Mathematical models of practical problems usually are designed to fit into well-known existing theory. At the same time new theoretical frameworks have to cope with criticism for lacking in practical relevance. To avoid such criticism, new theoretical results should come bundled with suggestions for improved mathematical models offered by the widened theory. Delivering such a bundle is the objective of this article.

The fundamental aim is to solve ill-posed nonlinear equations $F(x) = y$ only having a noisy measurement z of the exact right-hand side y at hand. Ill-posedness, here, mainly concerns a non-continuous dependence of the solutions on the right-hand side; existence of a solution will be assumed and uniqueness is not of interest.

The ill-posedness in connection with noisy data forces us to search for approximate but stable solutions. One way for obtaining such stable approximate solutions is to minimize a Tikhonov-type functional $S(F(x), z) + \alpha\Omega(x)$ over x. Here, S is a fitting functional and Ω stabilizes the problem. The concrete setting will be made precise in Section 2.

*Department of Mathematics, Chemnitz University of Technology, 09107 Chemnitz, Germany, email: jens.flemming@mathematik.tu-chemnitz.de

For ill-posed linear equations there is a comprehensive theory on regularization, especially on Tikhonov regularization, described, e.g., in [6]. First successful convergence rates results for ill-posed nonlinear equations in Hilbert spaces can be found in [7] and [6, Chap. 10]. Results on linear and nonlinear equations in Banach spaces are provided, e.g., in [2, 4, 13, 14, 20, 22, 24]. The cited publications, and almost all other publications on Tikhonov regularization, focus on Hilbert or Banach space norms as fitting functionals. But, mainly motivated by imaging applications (see, e.g., [1,19]), the interest in non-metric fitting functionals is growing. Here, "non-metric" means that at least one of the metric axioms, usually symmetry or the triangle inequality, is not satisfied by the fitting functional. Thus, there is the need to develop a theoretic framework for generalized Tikhonov regularization. Such Tikhonov-type approaches are also called variational regularization methods, emphasizing the demarcation to iterative methods.

In [25] and [17] quite general Tikhonov functionals are considered, but the authors focus on convergence theorems for a-posteriori parameter choices. We will present convergence rates results for a-priori parameter choices. A first result in this direction has been obtained in [21] and an extension of that convergence rates result has been given in [8]. The main question in deriving convergence rates for variational regularization with non-metric fitting functionals is, which properties the fitting functional has to satisfy. The assumptions posed in [21] and [8], particularly a triangle inequality, are too strong to apply the convergence rates results to practically relevant examples.

Therefore, in this paper we weaken the assumptions on the fitting functional, which allows us to provide examples of non-metric fitting functionals that on the one hand fit into the theory and on the other hand seem to be of practical relevance.

The paper has two main sections, the Sections 2 and 3. In Section 2 we describe a general theory for variational regularization with non-metric fitting functionals, i.e., we discuss the assumptions on the fitting functional and we provide theorems on existence, stability, and convergence, as well as on convergence rates. In Section 3 the theory is applied to two examples motivated by imaging applications. In the final section, Section 4, we summarize the results, discuss open questions and give an outlook on future work.

2 A general theory

In this section we present a general framework for handling variational regularization problems involving non-metric fitting functionals. At first we introduce notations and discuss two sets of assumptions on the fitting functional. Then theorems on existence, stability and convergence are given. The final subsection, containing the convergence rates theorem, is the main contribution of this paper.

2.1 Setting and assumptions

Since our approach does not require the involved spaces to have a uniform structure, i.e., to provide the notion of completeness, or to bring any operations, as addition or multiplication by scalars, with them, we work with topological spaces and need not assume that they are normed linear spaces or even Banach spaces. The notion of convergence is the only feature we use; and topological spaces exactly provide this feature, neither less nor more.

Let (X, τ_X), (Y, τ_Y), and (Z, τ_Z) be arbitrary topological spaces (convergence with respect to the topologies τ_X, τ_Y, and τ_Z will be denoted by "$\rightharpoonup$") and let $F : D(F) \subseteq X \to Y$ be any mapping from $D(F) \subseteq X$ into Y. For approximately solving the ill-posed equation

$$F(x) = y, \quad x \in D(F), \tag{2.1}$$

with right-hand side $y \in Y$ we consider the minimization of the Tikhonov-type functional

$$T_\alpha^z(x) := S(F(x), z) + \alpha\Omega(x), \quad x \in D(F), \tag{2.2}$$

over $D(F)$. Here $z \in Z$ represents a noisy measurement of the unknown right-hand side y. In this sense, X is the solution space, Y is the space of right-hand sides, and Z is the data space. The functional $S : Y \times Z \to [0, \infty]$ will be referred to as *fitting functional*, $\Omega : X \to (-\infty, \infty]$ will be referred to as *stabilizing functional*, and $\alpha \in (0, \infty)$ denotes the *regularization parameter*.

Throughout this section we assume that the following assumptions are satisfied.

Assumption 2.1. Assumptions on $F : D(F) \subseteq X \to Y$:

(i) F is sequentially continuous with respect to τ_X and τ_Y, i.e., $x_k \rightharpoonup x$ for $x, x_k \in D(F)$ implies $F(x_k) \rightharpoonup F(x)$.

(ii) $D(F)$ is sequentially closed with respect to τ_X, i.e., $x_k \rightharpoonup x$ for $x_k \in D(F)$ and $x \in X$ implies $x \in D(F)$.

Assumptions on $S : Y \times Z \to [0, \infty]$ (for arbitrary $y_k, y \in Y$ and $z_k, z \in Z$):

(iii) S is sequentially lower semi-continuous with respect to $\tau_Y \times \tau_Z$, i.e., if $y_k \rightharpoonup y$ and $z_k \rightharpoonup z$ then $S(y, z) \leq \liminf_{k\to\infty} S(y_k, z_k)$.

(iv) If $S(y, z_k) \to 0$ then there exists some $z \in Z$ such that $z_k \rightharpoonup z$.

(v) If $z_k \rightharpoonup z$ and $S(y, z) < \infty$ then $S(y, z_k) \to S(y, z)$.

Assumptions on $\Omega : X \to (-\infty, \infty]$:

(vi) Ω is sequentially lower semi-continuous with respect to τ_X, i.e., $x_k \rightharpoonup x$ for $x, x_k \in X$ implies $\Omega(x) \leq \liminf_{k\to\infty} \Omega(x_k)$.

(vii) The sets $M_\Omega(c) := \{x \in X : \Omega(x) \leq c\}$ are sequentially pre-compact with respect to τ_X for all $c \in \mathbb{R}$, i.e., each sequence in $M_\Omega(c)$ has a τ_X-convergent subsequence.

Note, that Assumption 2.1 allows Ω to attain negative values. But due to the pre-compactness of the sublevel sets $M_\Omega(c)$ and to the lower semi-continuity of Ω we know that Ω is bounded below.

Since in general $Y \neq Z$, the fitting functional S provides no direct way to check whether two elements $y_1 \in Y$ and $y_2 \in Y$ are equal. Thus, we have to introduce a notion of weak equality.

Definition 2.2. If for two elements $y_1, y_2 \in Y$ there exists some $z \in Z$ such that $S(y_1, z) = 0$ and $S(y_2, z) = 0$, we say that y_1 and y_2 are *S-equivalent with respect to z.* Further, we say that $x_1, x_2 \in X$ are S-equivalent if $F(x_1)$ and $F(x_2)$ are S-equivalent.

Assumption 2.1 can be simplified if $Y = Z$.

Proposition 2.3. *Assume that $Y = Z$ and that $S : Y \times Y \to [0, \infty]$ satisfies the following properties (for arbitrary $y_k, y, \tilde{y} \in Y$):*

(i) $S(y, \tilde{y}) = 0$ *if and only if* $y = \tilde{y}$.

(ii) S *is sequentially lower semi-continuous with respect to* $\tau_Y \times \tau_Y$.

(iii) $S(y, y_k) \to 0$ *implies* $y_k \rightharpoonup y$.

(iv) *If* $S(y, y_k) \to 0$ *and* $S(\tilde{y}, y) < \infty$ *then* $S(\tilde{y}, y_k) \to S(\tilde{y}, y)$.

Then S satisfies Assumption 2.1 if τ_Z is the topology induced by S, i.e., $y_k \overset{\tau_Z}{\rightharpoonup} y$ if and only if $S(y, y_k) \to 0$, and two elements $y_1, y_2 \in Y$ are S-equivalent if and only if $y_1 = y_2$.

Proof. Obviously the items of Assumption 2.1 concerning S are satisfied. If y_1, y_2 are S-equivalent with respect to some $z \in Z$ then the first property implies $y_1 = z = y_2$. The converse implication is trivially true (take $z := y_1$). □

The following example shows that our general framework covers the standard setting for variational regularization.

Example 2.4. Assume that X and Y are Banach spaces and that τ_X and τ_Y are the corresponding weak topologies. Then for $S(y_1, y_2) := \|y_1 - y_2\|^p$, $p > 0$, the assumptions of Proposition 2.3 are satisfied.

2.2 Existence, stability, convergence

In the sequel we make use of the notion of Ω-*minimizing solutions*, which is introduced by the following proposition.

Proposition 2.5. *If for $y \in Y$ there exists an element $\bar{x} \in D(F)$ with $F(\bar{x}) = y$ and $\Omega(\bar{x}) < \infty$ then there exists an Ω-minimizing solution of $F(x) = y$, i.e., there exists an element $x^\dagger \in D(F)$ which satisfies $\Omega(x^\dagger) = \inf\{\Omega(x) : x \in D(F),\ F(x) = y\}$.*

Proof. The assertion is a direct consequence of Assumption 2.1. □

The proofs of the following three theorems use standard techniques and are quite similar to the corresponding proofs given in [14] or [21]. Therefore we do not repeat them here.

Theorem 2.6 (existence)**.** *For all $z \in Z$ and all $\alpha > 0$ the minimization problem $T^z_\alpha(x) \to \min\limits_{x \in D(F)}$ has a solution.*

Theorem 2.7 (stability). *Fix $z \in Z$ and $\alpha \in (0,\infty)$ and assume that $(z_k)_{k\in\mathbb{N}}$ is a sequence in Z satisfying $z_k \rightharpoonup z$. Further assume that there exists an element $\bar{x} \in D(F)$ with $S(F(\bar{x}),z) < \infty$ and $\Omega(\bar{x}) < \infty$. Then each sequence $(x_k)_{k\in\mathbb{N}}$ with $x_k \in \operatorname{argmin}_{x\in D(F)} T_\alpha^{z_k}(x)$ has a τ_X-convergent subsequence and for sufficiently large k the elements x_k satisfy $T_\alpha^{z_k}(x_k) < \infty$. Each limit $\tilde{x}$ of a τ_X-convergent subsequence $(x_{k_l})_{l\in\mathbb{N}}$ is a minimizer of T_α^z and we have $T_\alpha^{z_{k_l}}(x_{k_l}) \to T_\alpha^z(\tilde{x})$, $\Omega(x_{k_l}) \to \Omega(\tilde{x})$, and thus also $S(F(x_{k_l}), z_{k_l}) \to S(F(\tilde{x}), z)$.*

Theorem 2.8 (convergence). *Assume that $y \in Y$, that $(z_k)_{k\in\mathbb{N}}$ is a sequence in Z satisfying $S(y,z_k) \to 0$, that $(\alpha_k)_{k\in\mathbb{N}}$ is a sequence in $(0,\infty)$ satisfying $\alpha_k \to 0$ and $\frac{S(y,z_k)}{\alpha_k} \to 0$, and that there exists an element $\bar{x} \in D(F)$ with $F(\bar{x}) = y$ and $\Omega(\bar{x}) < \infty$. Then each sequence $(x_k)_{k\in\mathbb{N}}$ with $x_k \in \operatorname{argmin}_{x\in D(F)} T_{\alpha_k}^{z_k}(x)$ has a τ_X-convergent subsequence and $T_{\alpha_k}^{z_k}(x_k) \to 0$. Each limit of a τ_X-convergent subsequence $(x_{k_l})_{l\in\mathbb{N}}$ is S-equivalent (cf. Definition 2.2) to each solution of $F(x) = y$. In addition, each such limit $\tilde{x}$ satisfies $\Omega(\tilde{x}) \leq \Omega(x^*)$ for all solutions x^* of $F(x) = y$. If $\tilde{x}$ is a solution then it is an Ω-minimizing solution and $\Omega(\tilde{x}) = \lim_{l\to\infty} \Omega(x_{k_l})$.*

2.3 Convergence rates

We consider equation (2.1) with a fixed right-hand side $y := y^0 \in Y$. By $x^\dagger \in X$ we denote one fixed Ω-minimizing solution of (2.1), where we assume that there exists a solution $\bar{x} \in D(F)$ with $\Omega(\bar{x}) < \infty$ (then Proposition 2.5 guarantees the existence of Ω-minimizing solutions).

Convergence rates results describe the dependence of the *solution error* on the *data error* if the data error is small. So at first we have to decide how to measure these errors. For this purpose we introduce a functional $D_{y^0} : Z \to [0,\infty]$ measuring the distance between the right-hand side y^0 and a data element $z \in Z$. On the solution space X we introduce a functional $E_{x^\dagger} : X \to [0,\infty]$ measuring the distance between the Ω-minimizing solution $x^\dagger$ and an approximate solution $x \in X$.

Because we are not interested in D_{y^0} at a fixed point, but we want to bound this functional, in the following by $z^\delta \in Z$ with $\delta \in (0,\infty)$ we denote an arbitrary data element satisfying

$$D_{y^0}(z^\delta) \leq \delta \tag{2.3}$$

(δ is called *noise level*). To guarantee the existence of some z^δ for each $\delta > 0$, we assume that there is some $z \in Z$ with $D_{y^0}(z) = 0$. A connection between S and D_{y^0} is established by the following assumption.

Assumption 2.9. There exists a monotonically increasing function $\psi : [0,\infty) \to [0,\infty)$ satisfying $\psi(\delta) = 0$ if and only if $\delta = 0$, $\psi(\delta) \to 0$ if $\delta \to 0$, and

$$S(y^0, z) \leq \psi(D_{y^0}(z)) \quad \text{for all } z \in Z \text{ with } D_{y^0}(z) < \infty. \tag{2.4}$$

This assumption provides the estimate

$$S(F(x^\dagger), z^\delta) = S(y^0, z^\delta) \leq \psi(D_{y^0}(z^\delta)) \leq \psi(\delta) < \infty. \tag{2.5}$$

If we take a sequence $(\delta_k)_{k\in\mathbb{N}}$ with $\delta_k \to 0$ we get $S(y^0, z^{\delta_k}) \le \psi(\delta_k) \to 0$. Thus, setting $z_k := z^{\delta_k}$, the assumptions of Theorem 2.8 are satisfied if we choose the sequence $(\alpha_k)_{k\in\mathbb{N}}$ in such a way that $\frac{\psi(\delta_k)}{\alpha_k} \to 0$ and $\alpha_k \to 0$.

Of course, we also need a connection between $E_{x^\dagger}$ and the terms of the Tikhonov functional (2.2) to prove convergence rates for $E_{x^\dagger}\big(x_{\alpha(\delta)}^{z^\delta}\big)$ with respect to δ, where $x_{\alpha(\delta)}^{z^\delta} \in \operatorname{argmin}_{D(F)} T_{\alpha(\delta)}^{z^\delta}$ and α depends on δ. This connection will be described by a *variational inequality* holding on a certain set $M \subseteq D(F)$. The technique of variational inequalities was introduced in [14] and has been extended in [13] and [2, 8, 21].

To formulate a variational inequality in our general setting we have to introduce an additional functional $S_{y^0} : Y \to [0,\infty]$, which measures the distance of some element $y \in Y$ to y^0. Note that using $S(y^0, \bullet) : Z \to [0,\infty]$ in sufficient conditions for convergence rates is quite difficult because such a condition then would depend on data elements $z \in Z$. A connection between S_{y^0} and S has to be given by

$$S_{y^0}(y) \le S(y,z) + S(y^0,z) \quad \text{for all } z \in Z \text{ and all } y \in Y. \tag{2.6}$$

As noted above, a variational inequality is connected to a set $M \subseteq D(F)$, on which the inequality shall hold. The only property such a set has to satisfy is that, given a parameter choice $\delta \mapsto \alpha(\delta)$,

$$\text{for all sufficiently small } \delta > 0 \text{ all minimizers } x_{\alpha(\delta)}^{z^\delta} \text{ belong to } M. \tag{2.7}$$

The following proposition provides an example for such a set (next to the trivial choice $M := D(F)$).

Proposition 2.10. *Let $\bar{\alpha} > 0$ and $\varrho > \Omega(x^\dagger)$. Further let $\delta \mapsto \alpha(\delta)$ be a parameter choice satisfying $\alpha(\delta) \to 0$ and $\frac{\psi(\delta)}{\alpha(\delta)} \to 0$ as $\delta \to 0$, and let $x_{\alpha(\delta)}^{z^\delta} \in \operatorname{argmin}_{D(F)} T_{\alpha(\delta)}^{z^\delta}$. Then there exists some $\bar{\delta} > 0$ such that*

$$x_{\alpha(\delta)}^{z^\delta} \in M := \{x \in D(F) : S_{y^0}(F(x)) + \bar{\alpha}\Omega(x) \le \varrho\bar{\alpha}\} \tag{2.8}$$

for all $\delta \in (0, \bar{\delta}]$.

Proof. The proof is similar to the corresponding one given in [8]. □

Definition 2.11. We say that the Ω-minimizing solution $x^\dagger$ satisfies a *variational inequality* if there exist some constant $\beta \in (0,\infty)$, a monotonically increasing function $\varphi : [0,\infty) \to [0,\infty)$ with $\varphi(0) = 0$, and a set $M \subseteq D(F)$ with property (2.7) such that $S_{y^0}(F(x)) < \infty$ for all $x \in M$ and

$$E_{x^\dagger}(x) \le \beta\big(\Omega(x) - \Omega(x^\dagger)\big) + \varphi\big(S_{y^0}(F(x))\big) \quad \text{for all } x \in M. \tag{2.9}$$

For future reference we formulate the following properties of the function φ appearing in Definition 2.11.

Assumption 2.12. The function $\varphi : [0,\infty) \to [0,\infty)$ satisfies:

(i) φ is monotonically increasing and $\varphi(0) = 0$;

(ii) there exists a constant $\gamma > 0$ such that φ is concave and strictly monotonically increasing on $[0,\gamma]$;

(iii) the inequality
$$\varphi(t) \leq \varphi(\gamma) + \left(\inf_{\tau\in[0,\gamma)} \frac{\varphi(\gamma)-\varphi(\tau)}{\gamma-\tau}\right)(t-\gamma)$$
is satisfied for all $t > \gamma$

If φ satisfies items (i) and (ii) of Assumption 2.12 and if φ is differentiable in γ, then item (iii) is equivalent to
$$\varphi(t) \leq \varphi(\gamma) + \varphi'(\gamma)(t-\gamma) \quad \text{for all } t > \gamma.$$
For example, $\varphi(t) = t^\mu$, $\mu \in (0,1]$, satisfies Assumption 2.12 for each $\gamma > 0$. The function
$$\varphi(t) = \begin{cases} (-\ln(t))^{-\mu}, & t \leq \mathrm{e}^{-\mu-1}, \\ \left(\frac{1}{\mu+1}\right)^{\mu} + \mu\left(\frac{\mathrm{e}}{\mu+1}\right)^{\mu+1}(t-\mathrm{e}^{-\mu-1}), & \text{else} \end{cases}$$
with $\mu > 0$ has a sharper cusp at zero than monomials and satisfies Assumption 2.12 for $\gamma \in (0, \mathrm{e}^{-\mu-1}]$.

Example 2.13. Consider the Banach space setting of Example 2.4, i.e., $Z = Y$ and $S(y_1, y_2) = \|y_1 - y_2\|^p$ for some $p \in (0,\infty)$. Setting $D_{y^0}(y) := \|y - y^0\|$ and $\psi(\delta) := \delta^p$ Assumption 2.9 is satisfied and estimate (2.5) coincides with the standard assumption $\|y^\delta - y^0\| \leq \delta$.

As error measure $E_{x^\dagger}$ in Banach spaces one usually uses the Bregman distance (cf. [4])
$$E_{x^\dagger}(x) := B_{\xi^\dagger}(x, x^\dagger) := \Omega(x) - \Omega(x^\dagger) - \langle \xi^\dagger, x - x^\dagger\rangle_{X^*,X}$$
with respect to some subgradient
$$\xi^\dagger \in \partial\Omega(x^\dagger) := \{\xi \in X^* : \Omega(x) \geq \Omega(x^\dagger) + \langle \xi, x - x^\dagger\rangle_{X^*,X}\}.$$
Other choices for $E_{x^\dagger}$ can be found in [3, Lemmas 4.4 and 4.6] and a whole class of very interesting error measures is proposed in [11].

For $S_{y^0}(y) := \max\{1, 2^{p-1}\}^{-1}\|y - y^0\|^p$ the triangle-type inequality (2.6) is satisfied and with $\varphi(t) := at^{\kappa/p}$ for $\kappa \in (0,p]$ and some $a > 0$ the variational inequality (2.9) attains the form
$$B_{\xi^\dagger}(x, x^\dagger) \leq \beta\big(\Omega(x) - \Omega(x^\dagger)\big) + c\|F(x) - F(x^\dagger)\|^\kappa$$
with $c \geq 0$. This variational inequality is equivalent to
$$-\langle \xi^\dagger, x - x^\dagger\rangle_{X^*,X} \leq \beta_1 B_{\xi^\dagger}(x, x^\dagger) + \beta_2\|F(x) - F(x^\dagger)\|^\kappa$$
for appropriate constants $\beta_1 < 1$ and $\beta_2 \geq 0$. The last inequality is of the type introduced in [13, 14].

The following main result of this paper is an adaption of Theorem 4.3 in [2] to our generalized setting. A similar result can be found in [11]. Unlike the corresponding proofs given in [2] and [11] our proof avoids the use of Young-type inequalities. Therefore it works also for non-differentiable functions φ in the variational inequality (2.9).

Theorem 2.14 (convergence rates)**.** *Let $x^\dagger$ satisfy a variational inequality (2.9) such that the associated function φ satisfies Assumption 2.12 and let $\delta \mapsto \alpha(\delta)$ be a parameter choice such that*

$$\inf_{\tau\in[0,\psi(\delta))} \frac{\varphi(\psi(\delta))-\varphi(\tau)}{\psi(\delta)-\tau} \geq \frac{\beta}{\alpha(\delta)} \geq \sup_{\tau\in(\psi(\delta),\gamma]} \frac{\varphi(\tau)-\varphi(\psi(\delta))}{\tau-\psi(\delta)}$$

with γ from Assumption 2.12. Then

$$E_{x^\dagger}\bigl(x_{\alpha(\delta)}^{z^\delta}\bigr) = \mathcal{O}\bigl(\varphi(\psi(\delta))\bigr) \quad \textit{if } \delta \to 0.$$

Remark 2.15. Note that by item (ii) of Assumption 2.12 a parameter choice as proposed in the theorem exists, i.e.,

$$\inf_{\tau\in[0,t)} \frac{\varphi(t)-\varphi(\tau)}{t-\tau} \geq \sup_{\tau\in(t,\gamma]} \frac{\varphi(\tau)-\varphi(t)}{\tau-t} > 0$$

for all $t \in (0,\gamma)$. One easily checks

$$\frac{\psi(\delta)}{\alpha(\delta)} \leq \frac{\varphi(\psi(\delta))}{\beta} \to 0 \quad \text{if } \delta \to 0$$

and if $\sup_{\tau\in(t,\gamma]} \frac{\varphi(\tau)-\varphi(t)}{\tau-t} \to \infty$ if $t \to 0$ then $\alpha(\delta) \to 0$ if $\delta \to 0$. Thus, the parameter choice satisfies the assumptions of Proposition 2.10.

If φ is differentiable in $(0,\gamma)$ then the proposed parameter choice is equivalent to

$$\alpha(\delta) = \frac{\beta}{\varphi'(\psi(\delta))}.$$

Proof. In the sequel we write α instead of $\alpha(\delta)$.

For sufficiently small $\delta > 0$, using $T_\alpha^{z^\delta}(x_\alpha^{z^\delta}) \leq T_\alpha^{z^\delta}(x^\dagger)$, (2.6), and (2.5), the variational inequality (2.9) implies

$$\begin{aligned} E_{x^\dagger}(x_\alpha^{z^\delta}) &\leq \tfrac{\beta}{\alpha}\bigl(T_\alpha^{z^\delta}(x_\alpha^{z^\delta}) - \alpha\Omega(x^\dagger) - S(F(x_\alpha^{z^\delta}),z^\delta)\bigr) + \varphi\bigl(S_{y^0}(F(x_\alpha^{z^\delta}))\bigr) \\ &\leq \tfrac{\beta}{\alpha}S(y^0,z^\delta) - \tfrac{\beta}{\alpha}S(F(x_\alpha^{z^\delta}),z^\delta) + \varphi\bigl(S(F(x_\alpha^{z^\delta}),z^\delta) + S(y^0,z^\delta)\bigr) \\ &\leq 2\tfrac{\beta}{\alpha}\psi(\delta) + \varphi\bigl(S(F(x_\alpha^{z^\delta}),z^\delta)+\psi(\delta)\bigr) - \tfrac{\beta}{\alpha}\bigl(S(F(x_\alpha^{z^\delta}),z^\delta)+\psi(\delta)\bigr) \end{aligned} \tag{2.10}$$

and therefore

$$E_{x^\dagger}\bigl(x_\alpha^{z^\delta}\bigr) \leq 2\tfrac{\beta}{\alpha}\psi(\delta) + \sup_{\tau\in[0,\infty)} \bigl(\varphi(\tau) - \tfrac{\beta}{\alpha}\tau\bigr).$$

If we can show, for sufficiently small δ and α as proposed in the theorem, that

$$\varphi(\tau) - \tfrac{\beta}{\alpha}\tau \leq \varphi(\psi(\delta)) - \tfrac{\beta}{\alpha}\psi(\delta) \quad \text{for all } \tau \geq 0, \tag{2.11}$$

then we obtain

$$\begin{aligned}
E_{x^\dagger}\left(x_\alpha^{z^\delta}\right) &\le \tfrac{\beta}{\alpha}\psi(\delta) + \varphi(\psi(\delta)) \le \psi(\delta) \inf_{\tau\in[0,\psi(\delta))} \frac{\varphi(\psi(\delta)) - \varphi(\tau)}{\psi(\delta) - \tau} + \varphi(\psi(\delta)) \\
&\le \psi(\delta)\frac{\varphi(\psi(\delta)) - \varphi(0)}{\psi(\delta) - 0} + \varphi(\psi(\delta)) = 2\varphi(\psi(\delta)).
\end{aligned}$$

Thus, it remains to show (2.11).

First we note that for fixed $t \in (0,\gamma)$ and all $\tau > \gamma$ item (iii) of Assumption 2.12 implies

$$\begin{aligned}
\frac{\varphi(\tau) - \varphi(t)}{\tau - t} &\le \frac{1}{\tau - t}\left(\varphi(\gamma) + \left(\inf_{\sigma\in[0,\gamma)} \frac{\varphi(\gamma) - \varphi(\sigma)}{\gamma - \sigma}\right)(\tau - \gamma) - \varphi(t)\right) \\
&\le \frac{1}{\tau - t}\left(\varphi(\gamma) + \frac{\varphi(\gamma) - \varphi(t)}{\gamma - t}(\tau - \gamma) - \varphi(t)\right) = \frac{\varphi(\gamma) - \varphi(t)}{\gamma - t}.
\end{aligned}$$

Using this estimate with $t = \psi(\delta)$ we can extend the supremum in the upper bound for α from $\tau \in (\psi(\delta),\gamma]$ to $\tau \in (\psi(\delta),\infty)$, that is,

$$\inf_{\tau\in[0,\psi(\delta))} \frac{\varphi(\psi(\delta)) - \varphi(\tau)}{\psi(\delta) - \tau} \ge \frac{\beta}{\alpha} \ge \sup_{\tau\in(\psi(\delta),\infty)} \frac{\varphi(\tau) - \varphi(\psi(\delta))}{\tau - \psi(\delta)}$$

or, equivalently,

$$\begin{aligned}
&\frac{\varphi(\psi(\delta)) - \varphi(\tau)}{\psi(\delta) - \tau} \ge \frac{\beta}{\alpha} \quad \text{for all } \tau \in [0,\psi(\delta)) \quad \text{and} \\
&\frac{\varphi(\tau) - \varphi(\psi(\delta))}{\tau - \psi(\delta)} \le \frac{\beta}{\alpha} \quad \text{for all } \tau \in (\psi(\delta),\infty)
\end{aligned}$$

These two inequalities together are equivalent to

$$\tfrac{\beta}{\alpha}(\psi(\delta) - \tau) \le \varphi(\psi(\delta)) - \varphi(\tau) \quad \text{for all } \tau \ge 0$$

and simple rearrangements yield (2.11). □

Remark 2.16. Let φ be differentiable in $(0,\infty)$ and continuous on $[0,\infty)$ with $\varphi'(t) \le c$ for some $c > 0$ and all $t > 0$, e.g. $\varphi(t) = at$ with a constant $a > 0$. Fixing $t > 0$, by the mean value theorem for each $\tau > 0$ there is some $\tilde{\tau} \in (\tau,t)$ such that $\varphi(t) - \varphi(\tau) = \varphi'(\tilde{\tau})(t - \tau)$. If we use $\varphi'(\tilde{\tau}) \le c$ and let τ tend to zero we get $\varphi(t) \le ct$ for all $t > 0$. From the proof of Theorem 2.14 we now see (cf. estimate (2.10)) that if we have a variational inequality with such a function φ then for any constant $\alpha^* \in \left(0, \frac{\beta}{c}\right)$ we get $E_{x^\dagger}\left(x_{\alpha^*}^{z^\delta}\right) = \mathcal{O}(\psi(\delta))$. This singular case corresponds to the exact penalization situation described in [4] for the Banach space setting of Example 2.13 with $\kappa = p = 1$.

3 Examples

To convince the reader of the necessity to investigate non-metric fitting functionals we give two illustrative examples. One exploiting all the features of the general theory described in Section 2, especially the distinction between the space Y of right-hand sides and the data space Z. And the other contenting itself with the $Y = Z$ setting of Proposition 2.3.

First, two fitting functionals are motivated in a more or less heuristic fashion. But in the second and third subsection we make their definitions precise and we show that the theorems of Section 2 apply. Variational inequalities are derived in the fourth subsection.

3.1 MAP estimation

An elegant way for motivating the minimization of various Tikhonov-like functionals comes from statistical inversion theory (see [18]). The basic idea is to regard all relevant variables, in our case x and z (y plays only a minor role), as random variables, here ξ and ζ, respectively, over a common probability space. Exploiting Bayes' formula for probability densities

$$p_\xi(x|\zeta = z) = \frac{p_\zeta(z|\xi = x)p_\xi(x)}{p_\zeta(z)}, \quad x \in X,\ z \in Z,\ p_\zeta(z) > 0, \tag{3.1}$$

and appropriately modelling the probability densities on its right-hand side, we seek for maximizers $x \in X$ of the conditional density $p_\xi(\bullet|\zeta = z)$ of ξ conditioned on $\zeta = z$. In other words: We maximize the probability that x is observed if z has been observed. Such a maximization problem can be reformulated as the minimization of a Tikhonov-like functional. The described approach often is refered to as *maximum a-posteriori probability estimation* (MAP estimation).

Note, that conditional probability densities of topological space valued random variables are a highly non-trivial topic (see [9]). Thus, at this point we have to be very careful in interpreting the terms in Bayes' formula (3.1). Since the MAP approach only shall motivate the examples we consider in this section, we avoid the technical details necessary to get reasoning bulletproof (the author has verified that a bulletproof version exists).

Now we have two tasks: modelling the densities $p_\zeta(z|\xi = x)$ and $p_\xi(x)$ (since $p_\zeta(z)$ is independent of x we do not care about this term), and deriving the corresponding minimization problem.

Consider the typical setting of imaging applications: We are interested in some quantity $x \in X$ which is not directly observable, but which can be made accessible through an energy density $y \in Y := L^1(T, \mu)$ arising from x via $y = F(x)$. Here $T \subseteq \mathbb{R}^2$ is a bounded set representing the surface of an image sensor, μ is a multiple (for scaling purposes) of the Lebesgue measure on T, and $L^1(T, \mu)$ is the Banach space of all real-valued μ-integrable functions. The image sensor, e.g. a CCD image sensor as used for ordinary digital cameras, does not provide us with this density function, but it counts the number of particles (e.g. photons) impinging on the mutually disjoint pixels $T_1, \dots, T_n \subseteq T$,

$n \in \mathbb{N}$, of the sensor during a fixed time interval. The number of particles arriving at T_i is, up to noise related fluctuations, determined by the energy density y on T_i. Instead of y, we only get a vector $z = (z_1, \dots, z_n) \in Z := \mathbb{N}_0^n$.

To model $p_\zeta(z|\xi = x)$ we have to make assumptions on the nature of the noise. Restricting our attention to applications where the number of impinging particles is very small (e.g. astronomy, medical imaging by PET or SPECT), it is sensible to assume that ζ_i follows a Poisson distribution with mean $\int_{T_i} F(x)\,\mathrm{d}\mu$. Additionally assuming that $\zeta_1, \dots, \zeta_n$ are mutually independent this Poisson approach leads to

$$p_\zeta(z|\xi = x) = \prod_{i=1}^{n} \frac{\left(\int_{T_i} F(x)\,\mathrm{d}\mu\right)^{z_i}}{z_i!} \exp\left(-\int_{T_i} F(x)\,\mathrm{d}\mu\right) \tag{3.2}$$

as a density with respect to the counting measure on $\mathbb{N}_0^n$. Here we assume that the $\int_{T_i} F(x)\,\mathrm{d}\mu$ are strictly positive and, for simplicity, that $D(F) = X$ is satisfied.

Furthermore, we have to model the probability density p_ξ, which allows us to incorporate a-priori knowledge about the solutions of $F(x) = y$ or to prescribe desired properties of the solutions. A widely used model (see, e.g., [1]) is $p_\xi(x) = c\exp(-\alpha\Omega(x))$ as a density with respect to a suitable measure on X. Here $c > 0$ is a normalizing factor, $\alpha > 0$ is a shape parameter, and $\Omega : X \to \mathbb{R}$ determines the basic structure of p_ξ.

Finally, we reformulate the maximization of

$$p_\xi(x|\zeta = z) = \frac{c\exp(-\alpha\Omega(x))}{p_\zeta(z)} \prod_{i=1}^{n} \left(\frac{\left(\int_{T_i} F(x)\,\mathrm{d}\mu\right)^{z_i}}{z_i!} \exp\left(-\int_{T_i} F(x)\,\mathrm{d}\mu\right)\right)$$

over $x \in X$ as the minimization of a Tikhonov-like functional. Taking the negative logarithm and adding $\ln c - p_\zeta(z) + \sum_{i=1}^{n}(z_i \ln z_i - z_i - \ln z_i!)$, which is independent of x, gives the equivalent minimization problem

$$\sum_{i=1}^{n} \left(z_i \ln \frac{z_i}{\int_{T_i} F(x)\,\mathrm{d}\mu} + \int_{T_i} F(x)\,\mathrm{d}\mu - z_i \right) + \alpha\Omega(x) \to \min_{x \in X}. \tag{3.3}$$

The functional in (3.3) serves as one of the two announced example for variational regularization with non-metric fitting functionals and will be refered to as the *semi-discrete model*.

If we had chosen $T = \{1, \dots, n\} \subseteq \mathbb{N}$ and μ to be the counting measure on this set then we would have $Y = L^1(T, \mu) = \mathbb{R}^n$ and $T_i := \{i\}$ would lead to $\int_{T_i} F(x)\,\mathrm{d}\mu = [F(x)]_i$. Thus the minimization problem would reduce to

$$\sum_{i=1}^{n} \left(z_i \ln \frac{z_i}{[F(x)]_i} + [F(x)]_i - z_i \right) + \alpha\Omega(x) \to \min_{x \in X}, \tag{3.4}$$

which motivates the infinite dimensional analogue

$$\int_T z \ln \frac{z}{F(x)} + F(x) - z\,\mathrm{d}\mu + \alpha\Omega(x) \to \min_{x \in X} \tag{3.5}$$

with $T \subseteq \mathbb{R}^2$ and $Y = Z = L^1(T, \mu)$. The functional in (3.5) is the other example we will discuss in detail. It will be refered to as *continuous model*.

3.2 Continuous model

At first we analyze the continuous model formally given in (3.5); the analysis of the semi-discrete model will make use of the notations and results of this subsection.

To make the definition of the fitting functional in (3.5) precise we have to handle the cases where the quotient or the logarithm are not defined. For this purpose we need two auxiliary functions f and g. Define $f : [0, \infty) \to \mathbb{R}$ by

$$f(u) := \begin{cases} u \ln u + 1 - u, & u \in (0, \infty), \\ 1, & u = 0 \end{cases} \tag{3.6}$$

and $g : [0, \infty) \times [0, \infty) \to (-\infty, \infty]$ by

$$g(u, v) := \begin{cases} u f\left(\frac{v}{u}\right), & u, v \in (0, \infty), \\ u, & u \in (0, \infty),\ v = 0, \\ \infty, & u = 0,\ v \in (0, \infty), \\ 0, & u = v = 0. \end{cases} \tag{3.7}$$

The function f is convex and continuous, and for $(\tilde{u}, \tilde{v}) \neq (0, 0)$ and $u, v \in (0, \infty)$ we have $u f\left(\frac{v}{u}\right) \to g(\tilde{u}, \tilde{v})$ if $(u, v) \to (\tilde{u}, \tilde{v})$.

Lemma 3.1. *The function g defined by (3.7) is nonnegative, $g(u, v) = 0$ if and only if $u = v$, g is convex, and g is lower semi-continuous.*

Proof. For $u, v \in (0, \infty)$ and $w := \frac{v}{u} \in (0, \infty)$ consider

$$u f\left(\frac{v}{u}\right) \geq 0 \quad \Leftrightarrow \quad f\left(\frac{v}{u}\right) \geq 0 \quad \Leftrightarrow \quad f(w) \geq 0 \quad \Leftrightarrow \quad w \ln w \geq w - 1.$$

Since the last inequality is satisfied for all $w \in (0, \infty)$ and equality holds if and only if $w = 1$, this proves the first two assertions. Convexity and lower semi-continuity follow from the fact that g can be expressed as a supremum of affine functions (see [12, p. 10] for details). □

Let $Y := Z := \{y \in L^1(T, \mu) : y \geq 0 \text{ a.e.}\}$ and for $0 \leq a < b < \infty$ set $Y_a^b := \{y \in L^1(T, \mu) : a \leq y \leq b \text{ a.e.}\}$. Further, let τ_Y be the topology on Y induced by the weak topology on $L^1(T, \mu)$. We define

$$S(y_1, y_2) := \begin{cases} \int_T g(y_1, y_2) \, \mathrm{d}\mu, & y_1, y_2 \in Y_a^b, \\ 0, & y_1 = y_2 \notin Y_a^b, \\ \infty, & \text{else}. \end{cases} \tag{3.8}$$

Since g is lower semi-continuous $g(y_1, y_2) : T \to [0, \infty]$ $(y_1, y_2 \in Y_a^b)$ is measurable and thus the integral in the definition of S exists (but may be infinite). If we neglect the bounds a, b and assume that y_1 and y_2 are probability densities then S is the Kullback-Leibler divergence. In this case the properties shown below are well known (see, e.g., [1,10,21]). For settings not restricted to probability densities the functional S is studied in [5,15,21,23].

We now show that S satisfies the assumptions of Proposition 2.3 and thus, it fulfills Assumption 2.1. That S is nonnegative and $S(y,\tilde{y}) = 0$ if and only if $y = \tilde{y}$ is a direct consequence of Lemma 3.1. The following propositions show that S also satisfies the other properties listed in Proposition 2.3.

Proposition 3.2. *For all $y_1, y_2 \in Y$ the functional S defined by (3.8) satisfies*

$$\|y_1 - y_2\|^2_{L^1(T,\mu)} \leq 4b\mu(T)S(y_1, y_2).$$

Proof. Let $y_1, y_2 \in Y_a^b$ (otherwise the assertion is trivially true). If $y_2 \neq y_1 = 0$ on a set $\hat{T} \subseteq T$ with $\mu(\hat{T}) > 0$ then $g(y_1, y_2) = \infty$ on $\hat{T}$ and thus $S(y_1, y_2) = \infty$, i.e., the assertion is true. So in the remaining part of the proof we assume $y_1 \neq 0$ almost everythere. Following the main ideas in [10] and choosing $\tilde{T} \subseteq T$ such that $y_1 \neq y_2$ a.e. on $\tilde{T}$ and $y_1 = y_2$ a.e. on $T \setminus \tilde{T}$ we have

$$\begin{aligned}
\int_T |y_1 - y_2| \,\mathrm{d}\mu &= \int_{\tilde{T}} \frac{|y_1 - y_2|}{\sqrt{g(y_1, y_2)}} \sqrt{g(y_1, y_2)} \,\mathrm{d}\mu \\
&\leq \sqrt{\int_{\tilde{T}} \frac{(y_1 - y_2)^2}{g(y_1, y_2)} \,\mathrm{d}\mu} \sqrt{\int_{\tilde{T}} g(y_1, y_2) \,\mathrm{d}\mu}.
\end{aligned}$$

Setting $h(u) := \frac{(u-1)^2}{(u \ln u + 1 - u)(u+1)} > 0$ and noticing that this function is bounded above by 2 for $u \in (0, \infty)$, we conclude

$$\begin{aligned}
&\|y_1 - y_2\|^2_{L(T,\mu)} \\
&\leq S(y_1, y_2) \int_{\tilde{T}} \frac{(y_1 - y_2)^2}{g(y_1, y_2)} \,\mathrm{d}\mu = S(y_1, y_2) \int_{\tilde{T}} \frac{\left(\frac{y_2}{y_1} - 1\right)^2}{f\left(\frac{y_2}{y_1}\right)\left(\frac{y_2}{y_1} + 1\right)} (y_1 + y_2) \,\mathrm{d}\mu \\
&\leq S(y_1, y_2) \int_T 2bh\left(\tfrac{y_2}{y_1}\right) \mathrm{d}\mu \leq 4b\mu(T)S(y_1, y_2).
\end{aligned}$$

□

Proposition 3.3. *Let S be defined as in (3.8) and let $y, y_k, \tilde{y} \in Y$ such that $S(y, y_k) \to 0$ and $S(\tilde{y}, y) < \infty$. With the additional assumption $a > 0$ we then have $S(\tilde{y}, y_k) \to S(\tilde{y}, y)$.*

Proof. For $y = \tilde{y}$ the asserion is trivial, so assume $y \neq \tilde{y}$. $S(\tilde{y}, y) < \infty$ implies $\tilde{y}, y \in Y_a^b$ and analogously $S(y, y_k) \to 0$ implies $y_k \in Y_a^b$ for sufficiently large k. Hence, we have

$$\begin{aligned}
|S(\tilde{y}, y_k) - S(\tilde{y}, y)| &= \left| \int_T y_k \ln \frac{y_k}{\tilde{y}} - y_k - y \ln \frac{y}{\tilde{y}} + y \,\mathrm{d}\mu \right| \\
&= \left| \int_T y_k \ln \frac{y_k}{y} + y - y_k + (y - y_k) \ln \frac{\tilde{y}}{y} \,\mathrm{d}\mu \right| \\
&\leq S(y, y_k) + \left(\ln \frac{b}{a}\right) \|y - y_k\|_{L^1(T,\mu)} \to 0
\end{aligned}$$

by Proposition 3.2. □

Proposition 3.4. *S defined by (3.8) is weakly lower semi-continuous on $Y \times Y$ and hence weakly sequentially lower semi-continuous.*

Proof. The proof is an adaption of the corresponding proofs in [21] and [12]. We have to show that the level-sets $M_S(c) := \{(y_1, y_2) \in Y \times Y : S(y_1, y_2) \leq c\}$ are weakly closed for all $c \in \mathbb{R}$. For $c < 0$ these sets are empty. For $c \geq 0$ we have $M_S(c) = \{(y, y) : y \in Y\} \cup M_G(c)$ with

$$M_G(c) := \left\{ (y_1, y_2) \in Y_a^b \times Y_a^b : \int_T g(y_1, y_2)\, \mathrm{d}\mu \leq c \right\}.$$

Because the limit of weakly convergent sequences in Banach spaces is uniquely determined one immediately sees that $\{(y, y) : y \in Y\}$ is weakly closed.

So it remains to show that the sets $M_G(c)$ are weakly closed, too. Since the $M_G(c)$ are convex (g is convex) and convex sets in Banach spaces are weakly closed if and only if they are closed it, suffices to show the closedness of $M_G(c)$ for $c \geq 0$. So let $(y_1^k, y_2^k) \in M_G(c)$ and $(y_1, y_2) \in Y \times Y$ such that $y_1^k \to y_1$ and $y_2^k \to y_2$ in $L^1(T, \mu)$. Then $y_1, y_2 \in Y_a^b$ and there exist subsequences $(y_1^{k_n})_{n\in\mathbb{N}}$ and $(y_2^{l_n})_{n\in\mathbb{N}}$ of (y_1^k) and (y_2^k) converging almost everywhere pointwise to y_1 and y_2, respectively. By the lower semi-continuity of g and Fatou's lemma we now get

$$\int_T g(y_1, y_2)\, \mathrm{d}\mu \leq \int_T \liminf_{n\to\infty} g(y_1^{k_n}, y_2^{l_n})\, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_T g(y_1^{k_n}, y_2^{l_n})\, \mathrm{d}\mu \leq c,$$

i.e., $(y_1, y_2) \in M_G(c)$. □

3.3 Semi-discrete model

For the semi-discrete example (3.3) we use the same setting as for the continuous one (i.e., Y, τ_Y, Y_a^b, f, and g are defined as in subsection 3.2), but we set $Z := \mathbb{R}_+^n = [0, \infty)^n$ and

$$S(y, z) := \begin{cases} \sum_{i=1}^n g\big(\int_{T_i} y\, \mathrm{d}\mu, z_i\big), & y \in Y_a^b, \\ \infty, & \text{else}. \end{cases} \tag{3.9}$$

The sets $T_1, \ldots, T_n \subseteq T$ are mutually disjoint measurable sets satisfying $\mu(T_i) > 0$ for $i = 1, \ldots, n$ and the topology τ_Z shall be the usual topology on $\mathbb{R}_+^n$. Our aim is to show that S satifies Assumption 2.1.

Proposition 3.5. *Let S be defined as in (3.9) and let $y \in Y$ and $z, z_k \in Z$ such that $z_k \rightharpoonup z$ and $S(y, z) < \infty$. With the additional assumption $a > 0$ we then have $S(y, z_k) \to S(y, z)$.*

Proof. $S(y, z) < \infty$ implies $y \in Y_a^b$. Hence, we have

$$|S(y, z_k) - S(y, z)| = \left| \sum_{i=1}^n [z_k]_i \ln \frac{[z_k]_i}{\int_{T_i} y\, \mathrm{d}\mu} - [z_k]_i - z_i \ln \frac{z_i}{\int_{T_i} y\, \mathrm{d}\mu} + z_i \right|$$

if all components of z and z_k are nonzero. And this expression goes to zero if $[z_k]_i \to z_i$. If for some k some of the $[z_k]_i$ or z_i are zero, the same can be shown. □

Proposition 3.6. *S defined by (3.9) is weakly lower semi-continuous on $Y \times Z$ and hence weakly sequentially lower semi-continuous.*

Proof. The proof is a simplified version of the proof of Proposition 3.4. □

Proposition 3.7. *Assume that S is defined by (3.9). If $S(y, z_k) \to 0$ for $y \in Y$ and $z_k \in Z$ then there exists some $z \in Z$ such that $z_k \rightharpoonup z$.*

Proof. $S(y, z_k) \to 0$ implies $S(y, z_k) < \infty$ for large k and thus $y \in Y_a^b$, i.e., $S(y, z_k) = \sum_{i=1}^n g(\int_{T_i} y \,\mathrm{d}\mu, [z_k]_i)$. Hence, defining $z_i := \int_{T_i} y \,\mathrm{d}\mu$ for $i = 1, \ldots, n$ we have $g(z_i, [z_k]_i) \to 0$. If, for fixed i, $z_i = 0$ then this convergence together with the definition of g implies $[z_k]_i = 0$ for all sufficiently large k. If $z_i > 0$ then $[z_k]_i > 0$ has to hold for all sufficiently large k, since if this would not be true there would exist a subsequence $([z_{k_n}]_i)_{n\in\mathbb{N}}$ with $[z_{k_n}]_i = 0$ and this would give $g(z_i, [z_{k_n}]_i) = z_i > 0$, which contradicts $g(z_i, [z_k]_i) \to 0$. So we have $g(z_i, [z_k]_i) = z_i f(\frac{[z_k]_i}{z_i}) \to 0$ for $i = 1, \ldots, n$ and this is the case if and only if $[z_k]_i \rightharpoonup z_i$. □

The following technical proposition will be of use in the next subsection.

Proposition 3.8. *For all $y \in Y$ with $y > 0$ a.e. and all $z \in Z$ the functional S defined by (3.9) satisfies*

$$\left\| y - \sum_{i=1}^n z_i \frac{y}{\int_{T_i} y \,\mathrm{d}\mu} \chi_{T_i} \right\|_{L^1(T,\mu)}^2 \leq 4b\mu(T) S(y, z),$$

where χ_{T_i} is one on T_i and zero on $T \setminus T_i$.

Proof. Only the case $y \in Y_a^b$ is of interest. Then, with $\tilde{y} := \sum_{i=1}^n \frac{z_i y}{\int_{T_i} y \,\mathrm{d}\mu} \chi_{T_i}$, by Proposition 3.2 and under the assumption $z_i > 0$ for $i = 1, \ldots, n$ we have

$$\begin{aligned} \|y - \tilde{y}\|_{L^1(T,\mu)}^2 &\leq 4b\mu(T) \int_T g(y, \tilde{y}) \,\mathrm{d}\mu = 4b\mu(T) \sum_{i=1}^n \int_{T_i} g(y, \tilde{y}) \,\mathrm{d}\mu \\ &= 4b\mu(T) \sum_{i=1}^n \int_{T_i} z_i \frac{y}{\int_{T_i} y \,\mathrm{d}\mu} \ln \frac{z_i}{\int_{T_i} y \,\mathrm{d}\mu} + y - z_i \frac{y}{\int_{T_i} y \,\mathrm{d}\mu} \,\mathrm{d}\mu \\ &= 4b\mu(T) \sum_{i=1}^n \left(z_i \ln \frac{z_i}{\int_{T_i} y \,\mathrm{d}\mu} + \int_{T_i} y \,\mathrm{d}\mu - z_i \right) = 4b\mu(T) S(y, z). \end{aligned}$$

If some of the z_i equal zero then the analog calculation with $g(y, \tilde{y}) = y$ on T_i would lead to the desired result. □

3.4 Variational inequalities

In this subsection we derive variational inequalities for both the continuous and the semi-discrete example by assuming that a source condition is satisfied. In fact, we only give a quite simple variant. But more sophisticated ways for obtaining variational inequalities are also applicable. For a detailed discussion on the interplay of source conditions and variational inequalities we refer to [16, 24].

Let (X, τ_X) be a topological vector space and let $Y := \{y \in L^1(T, \mu) : y \geq 0 \text{ a.e.}\}$ as in subsection 3.2. Further, assume that $F = A : D(A) \subseteq X \to Y$ is a linear operator with adjoint $A^* : Y^* \to X^*$, where $Y^* = L^\infty(T, \mu)$ and X^* is the set of all τ_X-continuous linear functionals on X (we assume $X^* \neq \emptyset$). By $\partial\Omega(\tilde{x}) := \{\xi \in X^* : \Omega(x) \geq \Omega(\tilde{x}) + \xi(x - \tilde{x})\}$ we denote the subdifferential of Ω at $\tilde{x} \in X$ and we define the Bregman distance with respect to Ω, $\tilde{x}$, and $\tilde{\xi} \in \partial\Omega(\tilde{x})$ by

$$B_{\tilde{\xi}}(x, \tilde{x}) := \Omega(x) - \Omega(\tilde{x}) - \tilde{\xi}(x - \tilde{x}), \quad x \in X. \tag{3.10}$$

In this subsection we assume that Ω is convex. Then $B_{\tilde{\xi}}(\bullet, \tilde{x}) \geq 0$.

The following proposition provides a preliminary form of a variational inequality.

Proposition 3.9. *Let $x^\dagger$ be an Ω-minimizing solution of $Ax = y^0$ such that there exists some $\xi^\dagger \in \partial\Omega(x^\dagger)$ satisfying $\xi^\dagger = A^*\eta$ for some $\eta \in Y^*$. Then*

$$B_{\xi^\dagger}(x, x^\dagger) \leq \Omega(x) - \Omega(x^\dagger) + \|\eta\|_{L^\infty} \|Ax - Ax^\dagger\|_{L^1} \tag{3.11}$$

holds for all $x \in D(A)$.

Proof. We have

$$-\xi^\dagger(x - x^\dagger) = -(A^*\eta)(x - x^\dagger) = -\eta(Ax - Ax^\dagger) \leq \|\eta\|_{L^\infty} \|Ax - Ax^\dagger\|_{L^1},$$

which proves the assertion. □

It remains to verify the triangle-type inequality (2.6) with the L^1-norm on its left side and the fitting functional S on the right side.

Proposition 3.10. *Let S be the continuous fitting functional defined by (3.8). Then*

$$\tfrac{1}{8b\mu(T)} \|y - y^0\|_{L^1}^2 \leq S(y, z) + S(y^0, z) \quad \text{for all } z \in Y.$$

Proof. For $z \in Y$ by the triangle inequality for norms and by Proposition 3.2 we have

$$\tfrac{1}{8b\mu(T)} \|y - y^0\|_{L^1}^2 \leq \tfrac{1}{4b\mu(T)} \big(\|y - z\|_{L^1}^2 + \|y^0 - z\|_{L^1}^2\big) \leq S(y, z) + S(y^0, z).$$

□

Proposition 3.11. *Let S be the semi-discrete fitting functional defined by (3.9). Then*

$$\tfrac{1}{8b\mu(T)} \|y - y^0\|_{L^1}^2 \leq S(y, z) + S(y^0, z) \quad \text{for all } z \in \mathbb{R}^n_+.$$

Proof. For $z = (z_1, \ldots, z_n) \in \mathbb{R}^n_+$ and $y \in Y$ set $\tilde{y} := \sum_{i=1}^n \frac{z_i y}{\int_{T_i} y \, \mathrm{d}\mu} \chi_{T_i}$, where χ_{T_i} is one on T_i and zero on $T \setminus T_i$. Then by the triangle inequality for norms and by Proposition 3.8 we have

$$\tfrac{1}{8b\mu(T)} \|y - y^0\|_{L^1}^2 \leq \tfrac{1}{4b\mu(T)} \big(\|y - \tilde{y}\|_{L^1}^2 + \|y^0 - \tilde{y}\|_{L^1}^2\big) \leq S(y, z) + S(y^0, z).$$

□

Thus, setting $E_{x^\dagger}(x) := B_{\xi^\dagger}(x,x^\dagger)$ and $S_{y^0}(y) := \frac{1}{8b\mu(T)}\|y-y_0\|_{L^1}^2$, for both the continuous and the semi-discrete example, by Proposition 3.9 the source condition $\xi^\dagger = A^*\eta$ implies a variational inequality

$$E_{x^\dagger}(x) \le \Omega(x) - \Omega(x^\dagger) + \sqrt{8b\mu(T)}\|\eta\|_{L^\infty} S_{y^0}(F(x))^{\frac{1}{2}} \tag{3.12}$$

for all $x \in D(A)$. For $D_{y^0}(z) := S(y^0, z)$, i.e., $S(y^0, z^\delta) \le \delta$, Theorem 2.14 now provides the convergence rate

$$B_{\xi^\dagger}\left(x^{z^\delta}_{\alpha(\delta)}, x^\dagger\right) = \mathcal{O}(\sqrt{\delta}) \tag{3.13}$$

for a parameter choice satisfying

$$c\sqrt{\frac{\delta}{2b\mu(T)}} \le \alpha(\delta) \le \sqrt{\frac{\delta}{2b\mu(T)}}$$

for some $c \in (0,1]$. In case of the continuous example, this rate also has been obtained in [1] for the symmetric Bregman distance.

4 Conclusions and open questions

We have seen that convergence rates results can be stated also for very general frameworks of variational regularization. Here, generality addresses different points: non-metric fitting functionals S, noise measures D_{y^0} which are only weakly connected with S, a wide range of convergence rates expressed by the function φ, and the freedom to choose an arbitrary error measure $E_{x^\dagger}$.

Former results on variational regularization with non-metric fitting functionals, in particular those presented in [8, 21], were based on the standard Banach space setting where S is a power of the norm. This led to unsatisfactory results concerning the assumptions imposed on S. Distinguishing between the space Y of right-hand sides and the data space Z we forced ourselves to develop new proofs instead of, exaggerating a bit, replacing all the norms in the Banach space proofs by S, as done before. This led to useful results as the examples have shown.

The next steps in the near future should include numerical experiments to compare the quality of solutions obtained with non-metric approaches to the quality of solutions obtained by standard Tikhonov regularization. And, since we have a new theoretic framework, we should look for further practical examples profiting from the extended theory.

As we have seen, the technique of variational inequalities for proving convergence rates turns out to be flexible enough to be applied to non-norm settings. Concerning variational inequalities there remain different open questions: In [16] and [8] it is shown that, using Bregman distances to measure the solution error, variational inequalities are limited to low-order convergence rates, i.e., there are higher convergence rates, provided by other techniques (see, e.g., [20]), that cannot be obtained via variational inequalities. Now, that we have the freedom to choose other error measures $E_{x^\dagger}$, e.g. $\|\bullet - x^\dagger\|^{3/2}$

instead of $\|\bullet - x^\dagger\|^2$ in Hilbert spaces, it is not clear whether the restriction to low-order rates persists further.

Another open questions concerns the existence of converse results, i.e., assertions about the validity of variational inequalities if convergence rates are known. A first attempt in this direction is presented in [16], but making a detour via source conditions. Is there a direct way from convergence rates to variational inequalities, at least in some special cases?

Deducing source conditions from variational inequalities is a second step of the reverse direction. The interplay between source conditions and variational inequalities in Banach spaces is described in [16]. But, as stated there, answers to "converse questions" are missing for nonlinear operators. Thus, variational inequalities will remain an object of active research.

References

[1] M. Benning and M. Burger. Error estimation for variational models with non-Gaussian noise. Technical report, WWU Münster, 2009.

[2] R. I. Boţ and B. Hofmann. An extension of the variational inequality approach for nonlinear ill-posed problems. *Journal of Integral Equations and Applications*, 22, 2010. to appear.

[3] K. Bredies and D. A. Lorenz. Regularization with non-convex separable constraints. *Inverse Problems*, 25(8):085011, 14pp, 2009.

[4] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Problems*, (5):1411–1421, 2004.

[5] P. P. B. Eggermont. Maximum Entropy Regularization for Fredholm Integral Equations of the First Kind. *SIAM Journal on Mathematical Analysis*, 24(6):1557–1576, 1993.

[6] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems.* Number 375 in Mathematics and its Applications. Kluwer Academic Publishers, Dordrecht, 1996.

[7] H. W. Engl, K. Kunisch, and A. Neubauer. Convergence rates for Tikhonov regularization of non-linear ill-posed problems. *Inverse Problems*, 5(4):523–540, 1989.

[8] J. Flemming and B. Hofmann. A New Approach to Source Conditions in Regularization with General Residual Term. *Numerical Functional Analysis and Optimization*, 31(3):254–284, 2010.

[9] B. Fristedt and L. Gray. *A Modern Approach to Probability Theory.* Birkhäuser, 1997.

[10] G. L. Gilardoni. On Pinsker's Type Inequalities and Csiszár's f-divergences. Part I: Second and Fourth-Order Inequalities, 2008. Paper published electronically as arXiv:cs/0603097v2[cs.IT] under http://arxiv.org/abs/cs/0603097.

[11] M. Grasmair. Generalized Bregman Distances and Convergence Rates for Nonconvex Regularization Methods. Industrial Geometry report 104, University of Vienna, Vienna, Austria, May 2010.

[12] A. Gundel. *Robust Utility Maximization, f-Projections, and Risk Constraints.* PhD thesis, Humboldt-Universität zu Berlin, Germany, 2006.

[13] T. Hein and B. Hofmann. Approximate source conditions for nonlinear ill-posed problems – chances and limitations. *Inverse Problems*, 25(3):035003, 16pp, 2009.

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

[15] B. Hofmann and R. Krämer. On maximum entropy regularization for a specific inverse problem of option pricing. *Jounal of Inverse and Ill-posed Problems*, 13(1):41–63, 2005.

[16] B. Hofmann and M. Yamamoto. On the interplay of source conditions and variational inequalities for nonlinear ill-posed problems. *Applicable Analysis*, 89, 2010. to appear.

[17] B. Jin and J. Zou. Iterative Parameter Choice by Discrepancy Principle. Technical report 2010-02 (369), Department of Mathematics, Chinese University of Hong Kong, Shatin, Hong Kong, Feb 2010.

[18] J. Kaipio and E. Somersalo. *Computational and Statistical Inverse Problems.* Number 160 in Applied Mathematical Science. Springer, 2005.

[19] T. Le, R. Chartrand, and T. J. Asaki. A Variational Approach to Reconstructing Images Corrupted by Poisson Noise. *Journal of Mathematical Imaging and Vision*, 27(3):257–263, 2007.

[20] A. Neubauer, T. Hein, B. Hofmann, S. Kindermann, and U. Tautenhahn. Improved and extended results for enhanced convergence rates of Tikhonov regularization in Banach spaces. *Applicable Analysis*, 89, 2010. to appear.

[21] C. Pöschl. *Tikhonov Regularization with General Residual Term.* PhD thesis, University of Innsbruck, Austria, 2008.

[22] E. Resmerita. Regularization of ill-posed problems in Banach spaces: convergence rates. *Inverse Problems*, 21(4):1303–1314, 2005.

[23] E. Resmerita and R. S. Anderssen. Joint additive Kullback-Leibler residual minimization and regularization for linear inverse problems. *Mathematical Methods in the Applied Sciences*, 30(13):1527–1544, 2007.

[24] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging.* Number 167 in Applied Mathematical Sciences. Springer New York, 2009.

[25] A. N. Tikhonov, A. S. Leonov, and A. G. Yagola. *Nonlinear Ill-Posed Problems.* Number vol. 1. Chapman Hall, London, 1998.