Discretization of variational regularization in Banach spaces

Christiane Pöschl [*], Elena Resmerita [†], and Otmar Scherzer [‡]

September 22, 2010

Abstract

Consider a nonlinear ill-posed operator equation $F(u) = y$ where F is defined on a Banach space X. In this paper we analyze finite dimensional variational regularization, which takes into account operator approximations and noisy data. As shown in the literature, depending on the setting, convergence of the regularized solutions of the finite dimensional problems can be with respect to the strong, or just a weak topology. In this paper our contribution is twofold. First, we derive convergence rates in terms of Bregman distances in the convex regularization setting under appropriate sourcewise representation of a solution of the equation. Secondly, for particular regularization realizations in nonseparable Banach spaces, we discuss finite dimensional approximations of the spaces and the type of convergence, which is needed in the convergence analysis. These considerations lay the fundament for efficient numerical implementation. In particular, we emphasize on the space X of finite total variation functions and analyze in detail the cases when X is the space of functions of finite bounded deformation and the L^∞–space. The latter two settings are of interest in numerous problems arising in optimal control, machine learning and engineering.

Key words: Ill-posed problem, Regularization, Bregman Distance, Strict Convergence

1 Introduction

Let $F : X \to Y$ be a nonlinear operator with domain $\mathcal{D}(F)$, where X is a Banach space and Y is a Hilbert space. We would like to approximate solutions

[*]Dept. of Information & Communication Technologies Universitat Pompeu Fabra C/ Tánger 122-140, 08018 Barcelona, Spain, *Christiane.Poeschl@univie.ac.at*

[†]Department of Industrial Mathematics, Johannes Kepler University, Altenbergerstraße 69, A-4040 Linz, Austria, *elena.resmerita@jku.at*

[‡]Computational Science Center, University of Vienna, Nordbergstraße 15, A-1090 Vienna, Austria and Radon Institute of Computational and Applied Mathematics, Altenbergerstraße 69, A-4040 Linz, Austria *otmar.scherzer@univie.ac.at*

of the ill-posed equation

$$F(u) = y \tag{1}$$

via variational regularization.

Let $\mathcal{R} : X \to [0, +\infty]$ be a penalty functional with nonempty domain $\mathcal{D}(\mathcal{R})$. An element $\bar{u} \in \mathcal{D}(F) \cap \mathcal{D}(\mathcal{R})$ is called an *$\mathcal{R}$-minimizing solution of* (1) if it solves the constraint optimization problem

$$\min \mathcal{R}(u) \text{ subject to } F(u) = y. \tag{2}$$

We assume that noisy data y^δ are given such that

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

In order to solve equation (1) numerically, the space X has to be approximated by a sequence of finite dimensional subspaces X_n. The situation when the spaces X and Y are Hilbert and the regularization is quadratic has been analyzed in [30], [32] (linear problems) and [31],[33] (nonlinear problems). However, recent advances in variational regularization theory deal with general Banach spaces. Although convergence of regularization methods in the general setting has been established (see, e.g., [40, Section 2], [35], [22]), the literature on analysis of discretization, and therefore numerical implementation, is not quite developed. The focus of this work is on problems modeled in nonseparable Banach spaces. In comparison with the Hilbert space theory, a significant complication is due to the fact that such a space cannot be approximated by a nested sequences of finite dimensional subspaces, with respect to the norm topology. We have in mind the space of bounded variation functions BV, the space of bounded deformation functions BD, and the space of essentially bounded functions L^∞ which are all not separable. However, as mentioned in [3, page 121] in the context of BV, "the norm topology is too strong for many applications", and in particular also when considering finite dimensional approximations. The reader is referred to [40, Section 4] for convergence results of finite dimensional regularization in general topological spaces and to, e.g, [17], [9], [8] for the BV case. More precisely, convergence was shown with respect to a weak topology of the space and with respect to the penalty function (see Section 2 for details). As explained in [6], Bregman distances are appropriate "measures" for quantifying convergence rate. Thus, deriving convergence rates (in the sense of Bregman distances) for variational regularization by discretization under appropriate sourcewise representation of a solution of the equation is our first concern. As stated in [40, Section 4], [17] and [8] it is natural to consider convergence of finite dimensional discretizations with respect to a topology which is weaker than the norm topology. Instead of the norm topology on X (as in the separable Hilbert space or in the separable Banach space setting), one uses a metric on X, which is induced from the metric of a superspace of X. As we show, in some examples, the convergence in the sense of such a metric is equivalent to approximation properties of the finite dimensional subspaces with respect to the weak topology. Thus, our second contribution is studying the convergence type for finite dimensional regularization. We discuss the case when X is the space of finite total variation functions (see [17], [9], [8]) from the perspective described above and analyze in detail the case when X is the space of functions of finite bounded deformation, and the L^∞–space. Needless to say that the BV-space and the BD-space are successfully employed in the theory of image processing and of elasto-plastic body deformations. Modeling regularization in the L^∞–space has not been widely studied in the inverse problems community. However, it is relevant in optimal control, signal processing, machine learning - see, e.g., [16], [20], [21], [24], [41], [14].

The paper is organized as follows. Section 2 specifies the assumptions, discusses well-posedness and convergence of the discretized regularization method in the case that the Banach space X is nonseparable, not even requiring that the regularization term is a convex function. Also, convergence rates with respect to Bregman distances are obtained in the convex regularization setting, under a standard source condition. The finite dimensional approximation of solutions of the equation $F(x) = y$ in separable Banach spaces X is briefly discussed in Section 3. Section 4 studies in detail the discretization of several nonseparable Banach spaces, such as the space of bounded variation functions, the space of bounded deformation functions and the space of essentially bounded functions. The inverse ground water filtration problem is analyzed in the natural setting of $L^\infty(\Omega)$, in Section 5.

2 The case of nonseparable Banach spaces

2.1 Main assumptions

Let X be a not necessarily separable Banach space which can be embedded into a separable Banach space Z. Let τ be a topology on X which is weaker than the norm topology on X. Therefore, we refer to τ as the *weak topology*. In addition let $\mathcal{R} : X \to [0, +\infty]$ be a proper functional.

We define a metric on the space X by the norm of Z induced on X and a pseudometric generated by the function $\mathcal{R}$:

$$d(u, v) = \|u - v\|_Z + |\mathcal{R}(u) - \mathcal{R}(v)|. \tag{4}$$

We shall denote by τ_d the topology generated by this metric. Relating to the above discussion, this is an intermediary topology between the norm topology on X and the weak topology τ.

The following elementary result gives us a motivation for approximating a nonseparable Banach space in a topology that is weaker than the norm topology.

Proposition 2.1. *A Banach space X is separable if and only if there exists a nested (increasing) sequence of finite dimensional subspaces $\{X_n\}$ such that*

$$\overline{\cup_{n\in\mathbb{N}} X_n} = X,$$

where the closure is considered with respect to the norm topology of X.

Consider a sequence of nested, finite dimensional subspaces $\{X_n\}$ which satisfies

$$\overline{\cup_{n\in\mathbb{N}} X_n}^{d} = X; \tag{5}$$

That is, X is the closure of the reunion of the subspaces X_n with respect to the topology of the metric d. This property holds for many nonseparable Banach spaces - see several examples in Section 4.

We are given approximation operators F_m of F, which have the same domain $\mathcal{D}(F)$ as F. We assume that the operators F, F_m satisfy the following approximation property:

$$\|F(u) - F_m(u)\| \leq \rho_m \text{ for all } u \in \mathcal{D}(F) \cap \mathcal{D}(\mathcal{R}) \ . \tag{6}$$

Here, the constant ρ_m should only depend on m and satisfy

$$\lim_{m\to\infty} \rho_m = 0 \ .$$

Denote

$$D_n := X_n \cap \mathcal{D}(F) \cap \mathcal{D}(\mathcal{R}), n \in \mathbb{N},$$

and assume that the sets D_n are nonempty. We are interested in approximating $\mathcal{R}$-minimizing solutions of equation (1) by solutions $u_{m,n}^{\alpha,\delta} \in D_n$ of the problem

$$\min\left\{\left\|F_m(u) - y^\delta\right\|^2 + \alpha\mathcal{R}(u)\right\} \quad \text{subject to } \ u \in D_n. \tag{7}$$

In order to pursue the analysis, we make several (standard) assumptions on the spaces X, Y, the operator F, the functional $\mathcal{R}$ (see also ([35, 22, 36]), and on the approximations X_n, F_m as well:

Assumption 2.2.

1. *The Banach space X is provided with a topology τ such that*
   - *The topology τ_d is finer than the topology τ.*
   - *The norm topology is finer than the topology τ_d.*
2. *The domain $\mathcal{D}(F)$ is τ-closed.*
3. *The operator $F : \mathcal{D}(F) \subseteq X \to Y$ is sequentially τ-weakly closed. That is, $\{u_k\} \subset \mathcal{D}(F)$, $u_k \xrightarrow{\tau} u$ and $F(u_k) \xrightarrow{w} v$ imply $u \in \mathcal{D}(F)$ and $v = F(u)$. Moreover, the operator F is continuous from $\mathcal{D}(F) \subset Z$ with the norm topology to $(Y, \|\cdot\|)$*
4. *For every $m \in \mathbb{N}$, the operator F_m is sequentially τ-weakly closed.*
5. *The function $\mathcal{R}$ is sequentially τ - lower semi-continuous.*
6. *For every $M > 0$, $\alpha > 0$ and every $m, n \in \mathbb{N}$, the sets*
   $$\{u \in X_n : \|F(u)\|^2 + \alpha\mathcal{R}(u) \leq M\} \tag{8}$$
   are τ-sequentially relatively compact.
7. *For every $u \in X$, there exists some $u_n \in X_n$ such that $d(u_n, u) \to 0$ as $n \to \infty$.*

2.2 Well-posedness of the discretized problem

We emphasize again one of the main ideas in this work: When discretizing a nonseparable Banach space, one could work with topologies which are weaker than the original norm topology and which might be more natural than the latter. A well-posedness result for problem (7) can be stated now in this theoretical setting.

Proposition 2.3. *Let $m, n \in \mathbb{N}$ and $\alpha, \delta > 0$ be fixed. Moreover, let assumptions 2.2 and* (6), (3) *be satisfied.*

Then, for every $y^\delta \in Y$, there exists at least one minimizer u of (7).

Moreover, the minimizers of (7) *are stable with respect to the data y^δ in the following sense: if $\{y_k\}_{k\in\mathbb{N}}$ converges strongly to y^δ, then every sequence $\{u_k\}_{k\in\mathbb{N}}$ of minimizers of* (7) *where y^δ is replaced by y_k has a subsequence $\{u_l\}_{l\in\mathbb{N}}$ which converges with respect to the topology τ to a minimizer $\tilde{u}$ of* (7) *and such that $\{\mathcal{R}(u_l)\}_{l\in\mathbb{N}}$ converges to $\mathcal{R}(\tilde{u})$, as $l \to \infty$.*

The proof is analogous to the one for Theorem 3.23 in [36], which in turn generalizes a proof in [15] from Hilbert to Banach spaces. With a different notation, the result and its proof can also be found in [40]. The same applies to the convergence result stated below - see [40, Sections 2-4]. However, for the sake of completeness and connection to the main results in Subsection 2.3 we have included the statement and the proof here.

Theorem 2.4. *Let assumptions 2.2 be satisfied. Moreover, assume that:*

(i) Equation (1) has an $\mathcal{R}$-minimizing solution $\bar{u}$ in the interior of $\mathcal{D}(\mathcal{R}) \cap \mathcal{D}(F)$, considered in the norm topology;

(ii) $v_n \in \mathcal{D}(F)$ for n sufficiently large, where $v_n \in X_n$ and $d(v_n, \bar{u}) \to 0$ as $n \to \infty$;

(iii) The parameter $\alpha = \alpha(m, n, \delta)$ is such that $\alpha \to 0$, $\frac{\delta^2}{\alpha} \to 0$, $\frac{{\rho_m}^2}{\alpha} \to 0$ and

$$\frac{\|F(v_n) - y\|}{\sqrt{\alpha}} \to 0 \tag{9}$$

as $\delta \to 0$ and $m, n \to \infty$.

If (6), (3) also hold, then every sequence $\{u_k\}$, with $u_k := u_{m_k,n_k}^{\alpha_k,\delta_k}$ and $\alpha_k := \alpha(m_k, n_k, \delta_k)$ where $\delta_k \to 0$, $n_k \to \infty$, $m_k \to \infty$ as $k \to \infty$ and u_k is a solution of (7), has a subsequence $\{u_l\}$ which converges with respect to the topology τ to an $\mathcal{R}$-minimizing solution $\tilde{u}$ of equation (1) and such that $\{\mathcal{R}(u_l)\}_{l\in\mathbb{N}}$ converges to $\mathcal{R}(\tilde{u})$, as $l \to \infty$. Moreover, if $\bar{u}$ is the unique solution of (1), then the entire sequence $\{u_k\}$ converges to $\bar{u}$ in the sense of τ and $\mathcal{R}$ as above.

Proof. From (4) and Assumption 2.2, Item 7 it follows that $\mathcal{R}(v_n) \to \mathcal{R}(\bar{u})$ for $n \to \infty$. From the definition of $u_{m,n}^{\alpha,\delta}$, the estimate (3) and (6), it follows that

$$\begin{aligned} \|F_m(u_{m,n}^{\alpha,\delta}) - y^\delta\|^2 + \alpha\mathcal{R}(u_{m,n}^{\alpha,\delta}) &\leq \|F_m(v_n) - y^\delta\|^2 + \alpha\mathcal{R}(v_n) \\ \leq (\|F_m(v_n) - F(v_n)\| + \|F(v_n) - F(\bar{u})\| &+ \|F(\bar{u}) - y^\delta\|)^2 + \alpha\mathcal{R}(v_n) \\ &\leq (\rho_m + \|F(v_n) - y\| + \delta)^2 + \alpha\mathcal{R}(v_n). \end{aligned} \tag{10}$$

Therefore,

$$\mathcal{R}(u_{m,n}^{\alpha,\delta}) \le \frac{(\rho_m + \|F(v_n) - y\| + \delta)^2}{\alpha} + \mathcal{R}(v_n)\ .$$

Assumption (iii) now guarantees that

$$\limsup \mathcal{R}(u_{m,n}^{\alpha,\delta}) \le \mathcal{R}(\bar{u}). \tag{11}$$

Observe that $\lim_{n\to\infty} \|F(v_n) - F(\bar{u})\| = 0$ since F is continuous at $\bar{u}$ (compare Assumption 2.2 Item 3) and (i), (ii)). By (iii), the quantity $(\rho_m + \|F(v_n) - F(\bar{u})\| + \delta)^2$ in (10) converges to zero as $\delta \to 0$ and $m, n \to \infty$, and then

$$\lim \|F_m(u_{m,n}^{\alpha,\delta}) - y^\delta\| = 0\ .$$

Therefore, by applying again (6) and (3), we also have that

$$F(u_{m,n}^{\alpha,\delta}) \to y, \tag{12}$$

with respect to the norm of Y, as $\delta, \alpha \to 0$ and $m, n \to \infty$. Consider $\alpha_k := \alpha(m_k, n_k, \delta_k)$ and $u_k := u_{m_k,n_k}^{\alpha_k,\delta_k}$. Note that $\{\|F_m(u_k)\|^2 + \alpha_k \mathcal{R}(u_k)\}$ is bounded. Hence, by (6) and by the compactness hypothesis in Assumption 2.2, there exists a subsequence $\{u_j\}_{j\in\mathbb{N}}$ which is τ-convergent to some $\tilde{u} \in D_n$. Due to the lower semicontinuity of $\mathcal{R}$ and (11), we get

$$\mathcal{R}(\tilde{u}) \le \liminf_{j\to\infty} \mathcal{R}(u_j) \le \limsup_{j\to\infty} \mathcal{R}(u_j) \le \mathcal{R}(\bar{u}).$$

We also have that $F(\tilde{u}) = y$, as $j \to \infty$ due to (12) and Item 4 in Assumption 2.2. Therefore, $\tilde{u}$ is an $\mathcal{R}$ minimizing solution of equation (1) and

$$\mathcal{R}(\tilde{u}) = \lim_{j\to\infty} \mathcal{R}(u_j).$$

If the solution $\bar{u}$ is unique, the only limit point of $\{u_k\}$ with respect to τ and $\mathcal{R}$ is $\bar{u}$. □

Remark 2.5. *In some situations, convergence of a sequence $\{u_k\}$ to u with respect to the topology τ and such that $\mathcal{R}(u_k) \to \mathcal{R}(u)$ implies $d(u_k, u) \to 0$ as $k \to \infty$. This happens for BV and BD which are embedded into $Z = L^1$, and τ is chosen as the weak* topology, $\mathcal{R}$ is the total variation and total deformation seminorm, respectively - see Section 4.*

2.3 Convergence Rates

In this section, we establish error estimates for the approximation method we analyze, with respect to Bregman distances. To this end, assume throughout this section that the regularization functional $\mathcal{R}$ is convex.

Recall that *the Bregman distance* with respect to a possibly non-smooth convex functional $\mathcal{R}$ is defined by

$$D_{\mathcal{R}}(v, u) = \{D_{\mathcal{R}}^{\xi}(v, u) : \xi \in \partial\mathcal{R}(u) \ne \emptyset\}, \quad u, v \in \mathcal{D}(\mathcal{R}),$$

where

$$D_{\mathcal{R}}^{\xi}(v,u)=\mathcal{R}(v)-\mathcal{R}(u)-\langle\xi,v-u\rangle.$$

More information about Bregman distances and their role in optimization and inverse problems can be found in [34]. Error estimates for variational or iterative regularization of (1) by means of a non-quadratic penalty have been shown in [6, 34, 35, 22, 7, 18]. The Bregman distance $D_{\mathcal{R}}$ associated with $\mathcal{R}$ was naturally chosen as the measure of discrepancy between the error estimates.

We assume Frechet differentiability of the operator F around $\bar{u}$ which is considered to be in the interior of $\mathcal{D}(F)\cap\mathcal{D}(\mathcal{R})$; moreover, assume that its extension to the space Z is also Frechet differentiable around $\bar{u}$. In fact, our study is based on the following source-wise representation:

There exists $\omega\in Y$ such that

$$\xi=F'(\bar{u})^{*}\omega\in\partial\mathcal{R}(\bar{u}),\tag{13}$$

and on the following nonlinearity condition (see also [35]):

There exist $\varepsilon>0$ and $c>0$ such that

$$\|F(u)-F(\bar{u})-F'(\bar{u})(u-\bar{u})\|\leq cD_{\mathcal{R}}(u,\bar{u}),\tag{14}$$

is satisfied for all $u\in\mathcal{D}(F)\cap U_{\varepsilon}(\bar{u})$ with respect to the above subgradient ξ and such that

$$c\|\omega\|<1\ .\tag{15}$$

Here

$$D_{\mathcal{R}}(u,\bar{u})=\mathcal{R}(u)-\mathcal{R}(\bar{u})-\langle\xi,u-\bar{u}\rangle,$$

where $\xi\in\partial\mathcal{R}(\bar{u})$ satisfies (13). Let us denote by

$$\gamma_n:=\|F'(\bar{u})(v_n-\bar{u})\|,\tag{16}$$

$$\lambda_n:=D_{\mathcal{R}}(v_n,\bar{u}).\tag{17}$$

Here $\{v_n\}$ is a sequence as in Theorem 2.4.

In the following we derive a relation between the Bregman distance and the metric d at $\bar{u}$. Observe that

$$D_{\mathcal{R}}(v_n,\bar{u})=\mathcal{R}(v_n)-\mathcal{R}(\bar{u})-\langle\omega,F'(\bar{u})(v_n-\bar{u})\rangle.$$

One has $\mathcal{R}(v_n)-\mathcal{R}(\bar{u})\to 0$ since $d(v_n,\bar{u})\to 0$, as $n\to\infty$ (see (4)). Moreover, $F'(\bar{u})(v_n-\bar{u})\to 0$. As a consequence, $\lim_{n\to\infty}\lambda_n=0$. Thus, convergence with respect to the metric d is stronger than convergence with respect to the related Bregman distance.

Theorem 2.6. *Suppose that Assumption 2.2, the assumptions in the previous result, inequalities (3), (13) and (14) hold. Moreover, assume that $\rho_m=O(\delta+\lambda_n+\gamma_n)$, with λ_n, γ_n given by (17), (16). If $\alpha\sim\max\{\delta,\lambda_n,\gamma_n\}$, then*

$$D_{\mathcal{R}}(u_{m,n}^{\alpha,\delta},\bar{u})=O(\delta+\lambda_n+\gamma_n).\tag{18}$$

Proof. We have

$$\begin{aligned}
\|F_m(u_{m,n}^{\alpha,\delta}) - y^\delta\|^2 + \alpha\mathcal{R}(u_{m,n}^{\alpha,\delta}) &\le \|F_m(v_n) - y^\delta\|^2 + \alpha\mathcal{R}(v_n) \\
&\le (\|F_m(v_n) - F(v_n)\| + \|F(v_n) - F(\bar{u})\| + \|F(\bar{u}) - y^\delta\|)^2 + \alpha\mathcal{R}(v_n) \\
&\le (\rho_m + \|F(v_n) - F(\bar{u}) - F'(\bar{u})(v_n - \bar{u})\| + \|F'(\bar{u})(v_n - \bar{u})\| + \delta)^2 + \alpha\mathcal{R}(v_n) \\
&\le (\rho_m + c\lambda_n + \gamma_n + \delta)^2 + \alpha\mathcal{R}(v_n)\ .
\end{aligned}$$

Denote

$$\beta_n := (\rho_m + c\lambda_n + \gamma_n + \delta)^2. \tag{19}$$

We use (14) and get

$$\begin{aligned}
\|F_m(u_{m,n}^{\alpha,\delta}) - y^\delta\|^2 + \alpha D_{\mathcal{R}}(u_{m,n}^{\alpha,\delta}, \bar{u}) &\le \beta_n + \alpha\mathcal{R}(v_n) - \alpha\mathcal{R}(\bar{u}) - \alpha\langle \xi, u_{m,n}^{\alpha,\delta} - \bar{u}\rangle \\
&= \beta_n + \alpha D_{\mathcal{R}}(v_n, \bar{u}) - \alpha\langle \xi, u_{m,n}^{\alpha,\delta} - v_n\rangle \\
&= \beta_n + \alpha\lambda_n - \alpha\langle \omega, F'(\bar{u})(u_{m,n}^{\alpha,\delta} - v_n)\rangle \\
&= \beta_n + \alpha\lambda_n - \alpha\langle \omega, F'(\bar{u})(u_{m,n}^{\alpha,\delta} - \bar{u})\rangle \\
&\quad + \alpha\langle \omega, F'(\bar{u})(v_n - \bar{u})\rangle \\
&\le \beta_n + \alpha\lambda_n + \alpha c\|\omega\| D_{\mathcal{R}}(u_{m,n}^{\alpha,\delta}, \bar{u}) \\
&\quad + \alpha\|\omega\|\|F(u_{m,n}^{\alpha,\delta}) - F(\bar{u})\| + \alpha\|\omega\|\gamma_n.
\end{aligned}$$

Therefore,

$$\|F_m(u_{m,n}^{\alpha,\delta}) - y^\delta\|^2 + \alpha(1 - c\|\omega\|) D_{\mathcal{R}}(u_{m,n}^{\alpha,\delta}, \bar{u}) \le \beta_n + \alpha\lambda_n + \alpha\|\omega\|(\zeta_n + \gamma_n), \tag{20}$$

where $\zeta_n = \|F_m(u_{m,n}^{\alpha,\delta}) - y\|$. Due to (15), the term $\alpha(1 - c\|\omega\|) D_{\mathcal{R}}(u_{m,n}^{\alpha,\delta}, \bar{u})$ is non-negative. Therefore,

$$\|F_m(u_{m,n}^{\alpha,\delta}) - y^\delta\|^2 \le \beta_n + \alpha\lambda_n + \alpha\|\omega\|(\zeta_n + \gamma_n). \tag{21}$$

Using (3) we have

$$\zeta_n^2 \le \left(\|F_m(u_{m,n}^{\alpha,\delta}) - y^\delta\| + \|y^\delta - y\|\right)^2 \le 2\|F_m(u_{m,n}^{\alpha,\delta}) - y^\delta\|^2 + 2\delta^2.$$

This together with inequality (21) implies

$$\zeta_n^2 \le 2\beta_n + 2\alpha\lambda_n + 2\alpha\|\omega\|\zeta_n + 2\alpha\|\omega\|\gamma_n + 2\delta^2,$$

which yields

$$\zeta_n \le 2\alpha\|\omega\| + \left(2\alpha^2\|\omega\|^2 + 2\delta^2 + 2\beta_n + 2\alpha\lambda_n + 2\alpha\|\omega\|\gamma_n\right)^{1/2}. \tag{22}$$

From (20), it follows that

$$\alpha(1 - c\|\omega\|) D_{\mathcal{R}}(u_{m,n}^{\alpha,\delta}, \bar{u}) \le \beta_n + \alpha\lambda_n + \alpha\|\omega\|\zeta_n + \alpha\|\omega\|\gamma_n,$$

with ζ_n estimated above. Using (19) and taking $\alpha \sim \max\{\delta, \lambda_n, \gamma_n\}$ yield the above convergence rate. □

3 The case of separable Banach spaces

Let X be a separable Banach space. Consider a nested sequence of finite dimensional subspaces X_n, $n \in \mathbb{N}$, such that

$$\overline{\cup_{n\in\mathbb{N}} X_n} = X,$$

where the closure is considered with respect to the norm topology of X. By letting $Z = X$, a weak topology τ on X, a regularization functional $\mathcal{R} : X \to [0, +\infty]$ and using the assumptions employed for the results in Section 2, one obtains stability and convergence results similar to Proposition 2.3 and Theorem 2.4. Moreover, if $\mathcal{R}$ is convex, then convergence rates can also be established.

Remark 3.1. *In some situations, convergence of a sequence $\{u_k\}$ to u with respect to the topology τ and such that $\mathcal{R}(u_k) \to \mathcal{R}(u)$ implies $\|u_k - u\| \to 0$, as $k \to \infty$. This property is known as the Radon-Riesz or Kadec-Klee or H-property - see. for instance [29, Section 2.5]. Locally uniformly convex reflexive Banach spaces, in particular Hilbert spaces enjoy such a property when τ is chosen as the weak topology on X and $\mathcal{R} = \|\cdot\|^p$, with $p \in (1, +\infty)$. The same for L^p spaces, $W^{m,p}$ spaces, but also for L^1 when τ is the weak topology and $\mathcal{R}$ is the Shannon entropy (see [5]).*

Sparsity regularization

Let $\{\phi_i\}$ be an orthonormal basis of $L^2(\Omega)$. Denote by X the Banach space ℓ^2 which is identified with the functions with bounded ℓ^2 Fourier coefficients. Let X_n be the linear span of the first n Fourier modes.

For sparsity regularization one usually takes $\mathcal{R}(u) = \sum_i |u_i|$, where $u = \sum_i u_i \phi_i$. Thus, we consider the regularization method of minimizing the functional

$$u \to \|F(u) - y^\delta\|^2 + \alpha \sum_i |u_i| \,,$$

In this case, the topology τ on X is the weak topology of ℓ^2. The regularization results apply also in this setting.

Another example of regularization term which promotes sparsity is

$$\mathcal{R}(u) = \sum_i |u_i|^p, \quad p \in (0, 1). \tag{23}$$

This setting with $X = \ell^2$ and τ taken as the weak topology of ℓ^2 is also covered by the regularization theory analyzed in this work. Moreover, convergence of a sequence $\{u_k\}$ to u with respect to the topology τ and such that $\mathcal{R}(u_k) \to \mathcal{R}(u)$ implies convergence of $\{u_k\}$ to u relative to the quasinorm (23), as $k \to \infty$, cf. [19].

4 Particular nonseparable spaces

4.1 The bounded variation functions space

Recall that, for a bounded Lipschitz domain $\Omega \subset \mathbb{R}^N$ and for a given $N \in \mathbb{N}$, the space $BV(\Omega)$ of $L^1(\Omega)$-functions of bounded variation mapping Ω into $\mathbb{R}$ can be defined as the set of functions $w \in L^1(\Omega)$ such that the total variation of w is finite, that is,

$$\int_\Omega |Dw|_p = \sup\left\{ \int w(x)\psi(x)dx : \psi \in \mathcal{C}_c^1(\Omega), |\psi(x)|_{p'} \leq 1 \text{ for all } x \in \Omega \right\} < \infty.$$

Here, $|\cdot|_{p'}$ denotes the $l_{p'}$ vector norm where $p' = p/(p-1)$ is the conjugate exponent to p. In particular we are interested in the cases $p = 1, 2$, where $l_{p'} = l_\infty, l_2$. The case $p = 2$ corresponds to *isotropic* total variation.

Let us recall several properties of the space $BV(\Omega)$:

- It is the dual of a separable Banach space (see [3, Remark 3.12]) when provided with the norm

$$\|u\|_{BV} = \|u\|_{L^1} + \int_\Omega |Du|_p .$$

- The space $BV(\Omega)$ is continuously embedded in $L^r(\Omega)$, where $1 \leq r \leq \frac{N}{N-1}$.

We consider the setting $X = BV(\Omega)$, with τ being the weak* topology on $BV(\Omega)$, and $Z = L^1(\Omega)$.

The functional $\mathcal{R}$ is the total variation seminorm. Consider

$$d(u,v) = \|u - v\|_{L^1(\Omega)} + \left| \int_\Omega |Du|_p - \int_\Omega |Dv|_p \right| .$$

The metric d is the metric used also in [8], which gives the so-called *strict convergence*, according to [3]. Moreover, the strict convergence of a sequence $\{u_k\}$ to u is equivalent to convergence with respect to the topology τ together with $\int_\Omega |Du_k|_1 \to \int_\Omega |Du|_1$ as $k \to \infty$, since weak* convergence of a sequence $\{u_k\}$ to u in $BV(\Omega)$ is equivalent to boundedness of $\{\|u_k\|_{BV})\}$ together with convergence of $\{u_k\}$ to u in $L^1(\Omega)$ - see, e.g., Proposition 3.13 in [3].

The choice of the vector norm in the definition of the bounded variation seminorm is of special importance for approximating the BV space by subspaces consisting of piecewise constant functions. Assume that $\{\Omega_j\}$ is a decomposition of Ω, and consider the following finite dimensional spaces:

$$X_n = \left\{ u_n = \sum_{j=1}^n u^j \chi_{\Omega_j} : u^j \in \mathbb{R}, 1 \leq j \leq n \right\} .$$

When considering a partition of Ω into uniform parallelepipeds we can only guarantee the density assumption (5) when considering the l_1-norm in the definition of BV - see [9]. If one wants an isotropic behavior of the regularization term one has to consider the l_2 norm in the definition of the BV-seminorm. The problem is that in the case of uniform parallelepipeds (for instance pixels in imaging), there is no convergence with respect to this isotropic BV-seminorm. One has to consider a more general partition of the domain Ω, that allows to approximate level lines with any direction. One idea is to use an irregular triangulation. These observations have been made by [8] and [9], here we only state their main result, concerning anisotropic and isotropic total variation:

Theorem 4.1 ([8, 9])**.** *Let $\Omega \subset \mathbb{R}^n$ be a polygonal domain and let $h_0 > 0$.*

- *Given $u \in BV(\Omega)$, then there exists a family $\{t_h : 0 < h \leq h_0\}$ of **triangulations** of Ω such that the mesh-size of t_h is at most h, and functions $u_h \in \mathcal{A}_h^0$, where $\mathcal{A}_h^0$ denotes the space of piecewise constant functions corresponding to the triangulation t_h, such that*

$$\|u - u_h\|_{L^1} + \left| \int_\Omega |Du|_2 - \int_\Omega |Du_h|_2 \right| \to 0 \quad \text{as } h \to 0 \, .$$

- *Given $u \in BV(\Omega)$. Let $\{\Omega_j\}$ be a decomposition of Ω, into **parallelepipeds**, such that the maximal length of a parallelepiped is smaller then h. Then there exist functions $u_h \in V_h^0$, where V_h^0 denotes the space of piecewise constant functions corresponding to the partition $\{\Omega_i\}$, such that*

$$\|u - u_h\|_{L^1} + \left| \int_\Omega |Du|_1 - \int_\Omega |Du_h|_1 \right| \to 0 \quad \text{as } h \to 0 \, .$$

Another type of approximation of X is the 'J_μ-approximation property' employed in [17] which, adapted to our notation reads as follows:

X_n is a Φ_α-approximation of X if, for each $u \in X$, there exists a sequence $\{u_n\} \subset X_n$ such that $\|u - u_n\|_Z \to 0$ and $\Phi_\alpha(u_n) \to \Phi_\alpha(u)$ as $n \to \infty$, for any $\alpha \geq 0$, where Φ_α is given by

$$\Phi_\alpha(u) = \left\| F(u) - y^\delta \right\|^2 + \alpha \mathcal{R}(u) \, .$$

In [17] one aims at approximating minimizers of Φ_α (which depends on α) in X by minimizers of Φ_α in X_n for a fixed $\alpha > 0$, while our aim is to approximate solutions of the operator equation by minimizers of Φ_α in X_n when the regularization parameter α depends on the dimension n. Also, note that strong convergence in the BV-space is shown in [26, 27] for regularization with a functional J which is of Shannon entropy type.

We summarize the results of this section in an example

Example 4.2. *Let $X = BV(\Omega)$ with the weak* topology τ. Moreover, let $Z = L^1(\Omega)$ and let d be the metric*

$$d(u, v) = \|u - v\|_{L^1(\Omega)} + \left| \int_\Omega |Du| - \int_\Omega |Dv| \right| \, . \tag{24}$$

Let F_m and $F : \mathcal{D}(F) \subseteq BV(\Omega) \to L^2(\hat{\Omega})$ satisfy the related conditions in Assumption 2.2.

Then according to Theorem 4.1, for every $u \in BV(\Omega)$ there exists an approximating sequence of piecewise constant functions. Consequently, minimization of the discretized regularized problem is well–posed, stable, and convergent (cf. Proposition 2.3). The piecewise constant regularizers $\{umn\}$ approximate the $\mathcal{R}$-minimizing solution $\bar{u}$ in the sense of the metric (24).

4.2 The bounded deformation functions space

In the following let $\Omega = (0,1)^N$ the open unit cube. We choose the simple geometry not to be forced to take into account approximations of Ω, or irregular meshes, for the finite element method considered below.

The space $BD(\Omega)$ [39] of functions with **bounded deformation** in an open set $\Omega \subset \mathbb{R}^N$ is defined as the set of functions $\boldsymbol{u} = (u^1, \ldots, u^N) \in L^1(\Omega; \mathbb{R}^N)$ such that the symmetric distributional derivative

$$E_{ij}\boldsymbol{u} := \frac{1}{2}(D_i u^j + D_j u^i)$$

is a (matrix-valued) measure with finite total variation in Ω:

$$BD(\Omega) := \left\{ \boldsymbol{u} \in L^1(\Omega; \mathbb{R}^N), E_{ij}(\boldsymbol{u}) \in M_1(\Omega), i,j = 1, \ldots, N \right\},$$

where $M_1(\Omega)$ denotes the space of bounded measures. $BD(\Omega)$ is a nonseparable Banach space provided with the norm

$$\|\boldsymbol{u}\|_{BD} = \|\boldsymbol{u}\|_{L^1(\Omega;\mathbb{R}^N)} + \underbrace{\sum_{i,j}^{N} \int_\Omega |E_{ij}(\boldsymbol{u})|}_{=:\int_\Omega |E\boldsymbol{u}|} .$$

This space is strictly larger than the space of bounded variation functions $BV(\Omega; \mathbb{R}^N)$. It was introduced in [37] and [28] and has been widely considered in the literature (see [38, 39]) in connection with the mathematical theory of plasticity. Several interesting properties of the BD space are as follows: It is the dual of a separable Banach space (see [39]); the space $BD(\Omega)$ is continuously embedded in $L^p(\Omega; \mathbb{R}^N)$, where $1 \le p \le \frac{N}{N-1}$. In addition the space $BD(\Omega; \mathbb{R}^N)$ is not separable – if it would be, the space $BV(\Omega; \mathbb{R}^N)$, which is the subspace of $BD(\Omega; \mathbb{R}^N)$ where all components u^j, $j = 2, \ldots, N$ vanish, would be as well. However, this is not true as stated already in the previous example.

Let us return to the finite dimensional regularization framework we investigate in this work.

Consider the setting $X = BD(\Omega)$, τ the weak* topology on $BD(\Omega)$ and $Z = L^1(\Omega; \mathbb{R}^N)$. We associate with $n \in \mathbb{N}$ the discretization size $h_n := \frac{1}{n}$, use multi-indices $\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_N)$, and set $A := \{0, \ldots, n\}^N$.

We consider the finite-dimensional product space of piecewise linear splines

$$X_n := \left\{ \boldsymbol{u}_n = \sum_{\alpha\in A} \sum_{k=1}^{N} u^k_{\boldsymbol{\alpha}} \Delta\left(\frac{\boldsymbol{x}-\boldsymbol{\xi}_{\boldsymbol{\alpha}}}{h_n}\right) : u^k_{\boldsymbol{\alpha}} \in \mathbb{R} \right\}$$

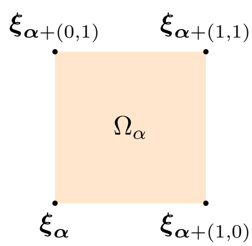
where $\boldsymbol{\xi}_{\boldsymbol{\alpha}} \in h_n \{0, 1, \ldots, n\}^N$, and Δ is the following function

$$\Delta(\boldsymbol{x}) := \prod_{i=1}^{N} \max(0, 1 - |x^i|) \ .$$

This finite element discretization has already been used in [10] for numerical minimization of variational energies. For $\boldsymbol{x} \in (0, h_n)^N$ the derivative of $\Delta((\boldsymbol{x} - \boldsymbol{\xi}_{\boldsymbol{\alpha}})/h_n)$ in direction $\boldsymbol{e}_j$ is given by

$$D_j\Delta\left(\frac{\boldsymbol{x}-\boldsymbol{\xi}_{\boldsymbol{\alpha}}}{h_n}\right) = sign(\boldsymbol{\xi}^j_{\boldsymbol{\alpha}} - x^j)\frac{1}{h_n}\prod_{i\neq j}^{N}\left(\max\left(0, 1 - \frac{|x^i - (\boldsymbol{\xi}_{\boldsymbol{\alpha}})^i|}{h_n}\right)\right) \ .$$

For every $\boldsymbol{\alpha}$ with $\alpha_i < n$ and for every $k, 1 \leq k \leq N$ define $A^k_{\boldsymbol{\alpha}} := \{\beta \in A_{\boldsymbol{\alpha}}, \boldsymbol{\beta}_k = \boldsymbol{\alpha}_k\}$. Additionally define Ω_α as the $N-$dimensional cube, spanned by the vectors $\{\boldsymbol{\xi}_{\boldsymbol{\alpha}} + \boldsymbol{e}_k h_n\}_{k=1\ldots N}$, and $A_{\boldsymbol{\alpha}} = \bigcup_{k=1}^N A^k_{\boldsymbol{\alpha}}$ (see Figure 1).



$$A_{\boldsymbol{\alpha}} = \{\boldsymbol{\alpha}, \boldsymbol{\alpha} + (0,1), \boldsymbol{\alpha} + (1,0), \boldsymbol{\alpha} + (1,1)\}$$
$$A^1_{\boldsymbol{\alpha}} = \{\boldsymbol{\alpha}, \boldsymbol{\alpha} + (0,1)\}$$
$$A^2_{\boldsymbol{\alpha}} = \{\boldsymbol{\alpha}, \boldsymbol{\alpha} + (1,0)\}$$

Figure 1: an example for the sets $A_{\boldsymbol{\alpha}}$

Moreover, for $\beta \in A_\alpha$ we have

$$\int_{\Omega_\alpha} D_j\Delta\left(\frac{\boldsymbol{x}-\boldsymbol{\xi}_\beta}{h_n}\right) d\boldsymbol{x} = \begin{cases} -\left(\frac{h_n}{2}\right)^{N-1} & \text{if } \beta^j = \alpha^j \\ +\left(\frac{h_n}{2}\right)^{N-1} & \text{if } \beta^j = \alpha^j + 1 \end{cases} \tag{25}$$

In the following we prove the main result on a pseudometric.

Theorem 4.3. *We assume that $h_n \to 0$ when $n \to \infty$. Then for every $\boldsymbol{u} \in BD(\Omega, \mathbb{R}^N) \cap L^r(\Omega, \mathbb{R}^N), 1 \leq r < \infty$, we can find a sequence $\{\boldsymbol{u}_n\}$, with $\boldsymbol{u}_n \in X_n$, such that*

$$\lim \int_\Omega |\boldsymbol{u} - \boldsymbol{u}_n|^r \, dx = 0 \ \textit{and} \ \lim_{n\to\infty} \int_\Omega |E\boldsymbol{u}_n| = \int_\Omega |E\boldsymbol{u}| \ .$$

Setting

$$d(\boldsymbol{u}_1, \boldsymbol{u}_2) = \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_{L^1(\Omega;\,\mathbb{R}^N)} + \left| \int_\Omega |E\boldsymbol{u}_1| \; - \int_\Omega |E\boldsymbol{u}_2| \right|, \tag{26}$$

we obtain $\lim_{n\to\infty} d(\boldsymbol{u}_n, \boldsymbol{u}) = 0$.

In order to prove this theorem, we need some additional facts on BD, given by the following Lemmas.

Lemma 4.4. *For every* $n \in \mathbb{N}$ *the inclusion* $X_n \subset BD(\Omega, \mathbb{R}^N)$ *holds and for each* $\boldsymbol{u} = (u^1, \ldots, u^N) \in X_n$,

$$\begin{aligned}\int_\Omega |E\boldsymbol{u}| \quad = \quad & \sum_\alpha \sum_{k=1}^N \left| \sum_{\boldsymbol{\beta}\in A_\alpha^k} (u^k_{\boldsymbol{\beta}+\boldsymbol{e}_k} - u^k_{\boldsymbol{\beta}}) \right| \left(\frac{h_n}{2}\right)^{N-1} \\ & + \sum_\alpha \sum_{k\neq l}^N \frac{1}{2} \left| \sum_{\boldsymbol{\beta}\in A_\alpha^l} (u^k_{\boldsymbol{\beta}+\boldsymbol{e}_l} - u^k_{\boldsymbol{\beta}}) + \sum_{\boldsymbol{\beta}\in A_\alpha^k} (u^l_{\boldsymbol{\beta}+\boldsymbol{e}_k} - u^l_{\boldsymbol{\beta}}) \right| \left(\frac{h_n}{2}\right)^{N-1}\end{aligned}$$

Proof. From the definition of $\boldsymbol{u} \in X_m$ we obtain

$$D_l u^k(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha}} u^k_{\boldsymbol{\alpha}} D_l \Delta \left(\frac{\boldsymbol{x} - \boldsymbol{\xi}_{\boldsymbol{\alpha}}}{h_n} \right) .$$

Moreover, since $\Omega = \bigcup_{\boldsymbol{\alpha}:\alpha_i<n} \Omega_{\boldsymbol{\alpha}}$, we can split up the integral. When integrating $D^l u_k + D^k u_l$ over Ω_α, we only have to sum over those $\boldsymbol{\beta}$ for which $\Delta\left(\frac{\boldsymbol{x}-\boldsymbol{\xi}_{\boldsymbol{\beta}}}{h_n}\right) \neq 0$. Hence we only sum over $\boldsymbol{\beta} \in A_\alpha$ in the inner sum.

$$\begin{aligned}\int_\Omega \left|D_l u^k + D_k u^l\right| &= \sum_{\boldsymbol{\alpha}} \int_{\Omega_{\boldsymbol{\alpha}}} \left|D_l u^k + D_k u^l\right| \\ &= \sum_{\boldsymbol{\alpha}} \int_{\Omega_\alpha} \left| \sum_{\boldsymbol{\beta}\in A_\alpha} \left(u^k_{\boldsymbol{\beta}} D_l \Delta \left(\frac{\boldsymbol{x}-\boldsymbol{\xi}_{\boldsymbol{\beta}}}{h_n}\right) + u^l_{\boldsymbol{\beta}} D_k \Delta \left(\frac{\boldsymbol{x}-\boldsymbol{\xi}_{\boldsymbol{\beta}}}{h_n}\right) \right) \right| .\end{aligned}$$

Next we use (25) and obtain

$$\int_{\Omega_\alpha} \left|D_l u^k + D_k u^l\right| = \left| \sum_{\boldsymbol{\beta}\in A_\alpha^l} (u^k_{\boldsymbol{\beta}+\boldsymbol{e}_l} - u^k_{\boldsymbol{\beta}}) + \sum_{\boldsymbol{\beta}\in A_\alpha^k} (u^l_{\boldsymbol{\beta}+\boldsymbol{e}_k} - u^l_{\boldsymbol{\beta}}) \right| \left(\frac{h_n}{2}\right)^{N-1}$$

The lemma follows from summing over all α and all $l, k = 1, \ldots, N$. ☐

Lemma 4.5. *1. If* $\{\boldsymbol{u}_n\} \subset BD(\Omega)$ *and* $\boldsymbol{u}_n \to \boldsymbol{u}$ *in* $(L^1(\Omega))^N$*, then*

$$\int_\Omega |E_{ij}\boldsymbol{u}| \leq \liminf_{n\to\infty} \int_\Omega |E_{ij}\boldsymbol{u}_n| \; .$$

2. *For every* $\boldsymbol{u} \in BD(\Omega) \cap L^r(\Omega), 1 \le r < \infty$, *there exists a sequence* $\{u_n\} \subset \mathcal{C}^{\infty}(\overline{\Omega})$ *such that*

$$\lim_{n\to\infty} \int |\boldsymbol{u} - \boldsymbol{u}_n|^r \, dx = 0 \qquad \lim_{n\to\infty} \int_\Omega |E_{ij}\boldsymbol{u}_n| = \int_\Omega |E_{ij}\boldsymbol{u}| \ .$$

Proof. 1. Follows from standard properties of convex measures.

2. See [38, Theoreme 3.2, Chapitre II].

□

Now we are ready for the proof of Theorem 4.3.

Theorem 4.3. The L^p-convergence of the picewise polinomial functions $\boldsymbol{u}_n$ can be found in the book of Ciarlet [12]. Due to Lemma 4.5, we can assume that $\boldsymbol{u} \in \mathcal{C}^\infty(\overline{\Omega})$. Set the coefficient of $\boldsymbol{u}_n$ as $(u_n)^k_{\boldsymbol{\alpha}} := u^k(\boldsymbol{\xi}_{\boldsymbol{\alpha}})$, then we have

$$\boldsymbol{u}_n = \sum_{k=1}^{N} \sum_{\boldsymbol{\alpha}} u^k(\boldsymbol{\xi}_{\boldsymbol{\alpha}}) \Delta \left(\frac{\boldsymbol{x} - \boldsymbol{\xi}_{\boldsymbol{\alpha}}}{h_n} \right) \ .$$

From Lemma 4.4 it follows that

$$\begin{aligned} \int_{\Omega_\alpha} |E_{kl}\boldsymbol{u}_n| &= \int_{\Omega_\alpha} \left| D_l u_n^k + D_k u_n^l \right| \\ &= \left| \sum_{\boldsymbol{\beta} \in A_\alpha^l} \left(u^k(\boldsymbol{\xi}_{\boldsymbol{\beta}+\boldsymbol{e}_l}) - u^k(\boldsymbol{\xi}_{\boldsymbol{\beta}}) \right) + \sum_{\boldsymbol{\beta} \in A_\alpha^k} \left(u^l(\boldsymbol{\xi}_{\boldsymbol{\beta}+\boldsymbol{e}_k}) - u^l(\boldsymbol{\xi}_{\boldsymbol{\beta}}) \right) \right| \left(\frac{h_n}{2} \right)^{N-1} \end{aligned} \tag{27}$$

Next we use the mean value theorem: For $m, n \in \{l, k\}, m \neq n$ we can find points $\boldsymbol{\eta}_{\boldsymbol{\beta},m,n}$ between $\boldsymbol{\xi}_{\boldsymbol{\beta}}$ and $\boldsymbol{\xi}_{\boldsymbol{\beta}+\boldsymbol{e}_m}$ such that

$$\begin{aligned} u^k(\boldsymbol{\xi}_{\boldsymbol{\beta}+\boldsymbol{e}_l}) - u^k(\boldsymbol{\xi}_{\boldsymbol{\beta}}) &= h_n D_l u^k(\boldsymbol{\eta}_{\boldsymbol{\beta},k,l}) \ , \\ u^l(\boldsymbol{\xi}_{\boldsymbol{\beta}+\boldsymbol{e}_k}) - u^l(\boldsymbol{\xi}_{\boldsymbol{\beta}}) &= h_n D_k u^l(\boldsymbol{\eta}_{\boldsymbol{\beta},l,k}) \ , \end{aligned}$$

hence from (27) it follows that

$$\int_{\Omega_\alpha} |E_{kl}\boldsymbol{u}_n| = \left| \sum_{\boldsymbol{\beta} \in A_\alpha^l} D_l u^k(\boldsymbol{\eta}_{\boldsymbol{\beta},k,l}) + \sum_{\boldsymbol{\beta} \in A_\alpha^k} D_k u^l(\boldsymbol{\eta}_{\boldsymbol{\beta},l,k}) \right| \underbrace{h_n^N}_{|\Omega_{\boldsymbol{\alpha}}|} \left(\frac{1}{2} \right)^{N-1}$$

Summing over all α and $l, k = 1, \dots N$ and taking the limit $h_n \to 0$ we have that

$$\lim_{n\to\infty} \int_\Omega |E_{ij}\boldsymbol{u}_n| = \int_\Omega |E_{ij}\boldsymbol{u}| \ .$$

□

We summarize the results of this section as follows:

Example 4.6. *Let $X = BD(\Omega)$ and τ the weak* topology. Moreover, let $Z = L^1(\Omega, \mathbb{R}^N)$ and let d be the metric*

$$d(\boldsymbol{u}, \boldsymbol{v}) = \|\boldsymbol{u} - \boldsymbol{v}\|_{L^1(\Omega, \mathbb{R}^N)} + \left| \int_\Omega |E\boldsymbol{u}| - \int_\Omega |E\boldsymbol{v}| \right| . \tag{28}$$

Let F_m and $F : \mathcal{D}(F) \subseteq BV(\Omega) \to L^2(\hat{\Omega})$ satisfy the related conditions in Assumption 2.2.

Take $h = 1/n$. Then according to Theorem 4.3, for every $\boldsymbol{u} \in BD(\Omega, \mathbb{R}^N) \cap L^r(\Omega, \mathbb{R}^N)$, $1 \le r < \infty$ there exists an approximating sequence of piecewise constant functions in the sense of the metric d. Consequently, minimization of the discretized regularized problem is well–posed, stable, and convergent according to Proposition 2.3. The piecewise constant regularizers $\{\boldsymbol{u}_{m,n}^{\alpha,\delta}\}$ approximate the $\mathcal{R}$-minimizing solution $\overline{\boldsymbol{u}}$ in the sense $\lim \boldsymbol{u}_{m,n}^{\alpha,\delta} = \overline{\boldsymbol{u}}$ in the weak star topology and $\lim \int_\Omega \left|E\boldsymbol{u}_{m,n}^{\alpha,\delta}\right| = \int_\Omega |E\overline{\boldsymbol{u}}|$.

4.3 The L^∞ space

In this section we analyze the following regularization method

$$u \to \|F(u) - y^\delta\|^2 + \alpha\|u\|_\infty .$$

In the sequel we show that there exist finite dimensional subspaces $\{X_n\}$ of $L^\infty(\Omega)$ satisfying equality (5). More precisely, there exist finite dimensional subspaces $\{X_n\}$ of $L^\infty(\Omega)$ such that for any $u \in L^\infty(\Omega)$ one can find $u_n \in X_n$, $n \in \mathbb{N}$ with

$$\lim_{n\to\infty} (\|u_n - u\|_{L^p} + |\|u_n\|_\infty - \|u\|_\infty|) = 0, \quad p \in [1, +\infty). \tag{29}$$

It is known that $L^\infty(\Omega)$ is not separable, while $L^p(\Omega)$, for every $p \in [1, +\infty)$, is separable. However, every function in $L^\infty(\Omega)$ can be approximated uniformly and thus, in the $L^\infty(\Omega)$ - norm by a sequence of simple functions. The proof of this statement is constructive. However, it provides a nonlinear approximation, in the sense that the piecewise constant functions which approximate the $L^\infty(\Omega)$ function do not yield linear subspaces - see, e.g., [13, Section 3.2]. This does not fit the theoretical framework we consider here. An alternative is to consider approximations of $L^\infty(\Omega)$ functions by piecewise constant functions in a weaker topology, as shown in the sequel.

For the sake of simplicity let $\Omega = (0, 1)^N$. Assume that $\{\Omega_j\}$ is a decomposition of Ω in parallelepipeds with maximal diagonal length h_n as in [9], and consider the following finite dimensional subspaces of $L^\infty(\Omega)$:

$$X_n = \left\{ u_n = \sum_{j=1}^n u^j \chi_{\Omega_j} : u^j \in \mathbb{R}, 1 \le j \le n \right\} .$$

The following results are essential in proving the main statement of this section:

Denote by $J_\epsilon * u$ the mollification of u, for every $\epsilon > 0$.

Theorem 4.7. *[p. 36 [2]] Let u be a function which is defined on $\mathbb{R}^N$ and vanishes identically outside Ω.*

If $u \in L^p(\Omega)$, then $J_\epsilon * u \in C^\infty(\mathbb{R}^N)$, in fact $J_\epsilon * u \in C^\infty(\overline{\Omega})$ and $J_\epsilon * u \in L^p(\Omega)$, for every $p \in [1, +\infty)$. Also,

$$\lim_{\epsilon \to 0_+} \|J_\epsilon * u - u\|_p = 0, \tag{30}$$

$$\|J_\epsilon * u\|_p \leq \|u\|_p, \quad \textit{for all } \epsilon > 0. \tag{31}$$

Lemma 4.8. *For every $n \in \mathbb{N}$ one has $X_n \subset L^\infty(\Omega)$ and for each $u_n \in X_n$*

$$\|u_n\|_\infty = \max_{1 \leq j \leq n} |u^j|.$$

Proof. Follows immediately from the definition of u_n, taking into account that $\mu(\Omega_j) > 0$, for $1 \leq j \leq n$. □

Lemma 4.9. *If $\{u_n\} \subset L^\infty(\Omega)$ and $u \in L^\infty(\Omega)$ are such that $\|u - u_n\|_{L^p} \to 0$ as $n \to \infty$, for some $p \in [1, +\infty)$ and $\|u_n\|_\infty \leq \|u\|_\infty$ for every $n \in \mathbb{N}$, then $\{u_n\}$ converges weakly* to u and $\|u_n\|_\infty \to \|u\|_\infty$ as $n \to \infty$.*

Proof. Since $\{\|u_n\|_\infty\}$ is bounded, there exists a subsequence $\{u_k\}$ which converges weakly* to some $v \in L^\infty(\Omega)$, cf. Alaoglu-Bourbaki Theorem, [23, p. 70]. By the definition of weak* convergence, one obtains that $\{u_k\}$ converges also weakly, with respect to $L^p(\Omega)$ to some v. Therefore $u = v$. In fact, every subsequence of $\{u_n\}$ converges weakly* to u, which yields weak* convergence of the entire sequence $\{u_n\}$ to u. In addition, the weak* lower semi continuity of the L^∞-norm implies $\|u\|_\infty \leq \liminf_{n\to\infty} \|u_n\|_\infty$. Thus the assertions are proved. □

Note that the previous result can be relaxed by assuming only weak convergence of $\{u_n\}$ in $L^p(\Omega)$.

Lemma 4.10. *For every $u \in C^\infty(\bar{\Omega})$, there exists a sequence $\{u_n\}$ with $u_n \in X_n$ such that*

$$\lim_{n\to\infty} \left(\|u_n - u\|_{L^p} + |\|u_n\|_\infty - \|u\|_\infty|\right) = 0, \quad p \in [1, +\infty).$$

Proof. Let ξ_j be the gravity centers of the parallelepipeds Ω_j. Define

$$u_n = \sum_{j=1}^{n} u(\xi_j) \chi_{\Omega_j}.$$

Then $\|u_n - u\|_{L^p} \to 0$ - see, for instance the book of Ciarlet [12]. Moreover, by using Lemma 4.8,

$$\|u_n\|_\infty = \max_{1 \leq j \leq n} |u(\xi_j)| \leq \max_{x \in \bar{\Omega}} |u(x)| = \|u\|_\infty, \qquad \text{for all } n \in \mathbb{N},$$

where the last equality holds due to the continuity of u on $\bar{\Omega}$. Thus, Lemma 4.9 applies and yields $|\|u_n\|_\infty - \|u\|_\infty| \to 0$ as $n \to \infty$. □

Theorem 4.11. *Assume that $h_n \to 0$ when $n \to \infty$. Then for every $u \in L^\infty(\Omega)$ one can find $u_n \in X_n$ such that (29) holds.*

Proof. Let $u \in L^\infty(\Omega)$ and $p \in [1, +\infty)$. By Theorem 4.7, there exists $\{u_j\} \subset C^\infty(\mathbb{R}^N)$, in fact in $C^\infty(\bar{\Omega})$, with $\{u_j\} \subset L^p(\Omega)$ such that

$$\lim_{j\to\infty} \|u_j - u\|_p = 0, \quad \text{and} \quad \|u_j\|_p \le \|u\|_p, \qquad \text{for all } \; j \in \mathbb{N}.$$

By letting $p \to +\infty$ in the last inequality and using Theorem 2.8 on p. 25 in [1], one also has

$$\|u_j\|_\infty \le \|u\|_\infty, \qquad \text{for all } \; j \in \mathbb{N}.$$

Lemma 4.9 applies and yields $\|u_j\|_\infty \to \|u\|_\infty$ as $j \to \infty$.

By consequence, every function in $L^\infty(\Omega)$ can be approximated by functions from $C^\infty(\overline{\Omega})$ in the sense stated at (29). Since every function $v \in C^\infty(\bar{\Omega})$ can be approximated by $u_n \in X_n$ as in (29) due to Lemma 4.10, the conclusion follows immediately. ☐

Remark 4.12. *One can define the subspaces X_n in the previous theorem also by means of piecewise polynomial functions of degree no bigger than one in each variable, which are continuous on $\bar{\Omega}$. Moreover, one can employ n-simplices instead of parallelipipeds and piecewise linear functions, according to Remark 3.8 in [9].*

We summarize the results of this section as follows:

Example 4.13. *Let $X = L^\infty(\Omega)$ and τ the weak* topology. Moreover, let $Z = L^p(\Omega)$, $p \in (1, +\infty)$ and let d be the metric*

$$d(u, v) = \|u - v\|_{L^p(\Omega)} + |\|u\|_\infty - \|v\|_\infty| \,. \tag{32}$$

Let F_m and $F : \mathcal{D}(F) \subseteq BV(\Omega) \to L^2(\hat{\Omega})$ satisfy the related conditions in Assumption 2.2.

Take $h = 1/n$. Then according to Theorem 4.11, for every $u \in L^\infty(\Omega)$ there exists a sequence of piecewise constant functions which approximates u in the sense of metric d. Consequently, minimization of the discretized regularized problem is well–posed, stable, and convergent according to Proposition 2.3. The piecewise constant regularizers $\{u_{m,n}^{\alpha,\delta}\}$ approximate the $\mathcal{R}$-minimizing solution $\bar{u}$ in the sense $\lim u_{m,n}^{\alpha,\delta} = \bar{u}$ in the weak star topology and $\lim \|u_{m,n}^{\alpha,\delta}\|_\infty = \|\bar{u}\|_\infty$.

The convergence described in the previous sentence is usually weaker than the convergence with respect to the metric given by (32). However, in special situations it implies strong convergence with respect to any L^p norm with $p \in [1, +\infty)$, cf. Theorem 3 in [21]. The fact that the two types of convergence in discussion are not equivalent in general is shown by the following counterexample.

Consider the Rademacher functions $f_n : [0, 1] \to \{-1, 1\}$,

$$f_n(t) = (-1)^{i+1} \;\text{ if }\; x \in \left[\frac{i-1}{2^n}, \frac{i}{2^n}\right), \;\; 1 \le i \le 2^n.$$

This sequence converges weakly star to zero in $L^\infty([0,1])$, but not in the $L^1([0,1])$ norm. Moreover, consider $g_n : [0,2] \to \mathbb{R}$,

$$g_n(t) = f_n(t), \text{ if } t \in [0,1]$$

and $g_n(t) = 1$ for $t \in [1,2]$. Then $\{g_n\}$ converges weakly star to $\chi_{[1,2]}$, the characteristic function of $[1,2]$ in $L^\infty([0,2])$ but not in the $L^1([0,2])$ norm. Note that $\|g_n\|_\infty = \|\chi_{[1,2]}\|_\infty = 1$. Thus, $\{g_n\}$ is a sequence in $L^\infty([0,2])$ which converges weakly star to $\chi_{[1,2]}$ and such that $\lim_{n\to\infty} \|g_n\|_\infty = \|\chi_{[1,2]}\|_\infty$, but $\lim_{n\to\infty} \left(\left\|g_n - \chi_{[1,2]}\right\|_{L^1} + \left|\|g_n\|_\infty - \|\chi_{[1,2]}\|_\infty\right|\right) \neq 0$.

Remark 4.14. *Given a direct problem formulated in L^∞, is it worth to formulate the inverse problem in L^∞ or is it more appropriate to formulate the problem in L^2, where we can get L^2-approximations? L^2 approximations are quite advantageous because in the L^∞ case, one might not even get convergence with respect to the L^1-norm, as demonstrated by the above counterexample. However, it is sometimes desirable that the regularized solutions are guaranteed to belong also to the L^∞ space. Moreover, the extremal behavior of the solution can be evaluated by L^∞ regularization, since convergence of the L^∞-norm of the regularized solutions to the L^∞ norm of the true solution is achieved. Also, in some inverse problems high interest is given to estimating a linear functional of the solution rather than the solution - according to the mollifier idea in [4], which corresponds to the weak-star approximation of our L^∞ regularization results.*

5 The inverse ground water filtration problem

We consider the problem of recovering the diffusion coefficient in

$$-(au_x)_x = f \text{ in } \Omega,$$

$$u(0) = 0 = u(1),$$

with $f \in L^2[0,1]$. The operator F is defined as the parameter-to-solution mapping

$$F : \mathcal{D}(F) := \{a \in L^\infty[0,1] : a(x) \geq c > 0, \ a.e.\} \to L^2[0,1],$$

$$a \mapsto F(a) := u(a),$$

where $u(a)$ is the unique solution of the above equation and c is a constant. More details about this ill posed problem can be found in [25] and [11, Chapter 1]. Note that $\mathcal{D}(F)$ is a subset of the interior of the nonnegative cone of $L^\infty[0,1]$. In fact L^∞ is the natural function space when formulating this problem. However, to the best of our knowledge, previous literature dealing with the problem from the regularization viewpoint has usually considered $\mathcal{D}(F)$ in H^1 which is embedded in L^∞, mainly due to the Hilbert space setting which is enforced by using H^1. The operator F is Fréchet differentiable from L^∞ to L^2 - see [25]. Note that $\mathcal{D}(F)$ is closed and convex with respect to L^2, so it is weakly closed in L^2. This implies that $\mathcal{D}(F)$ is weakly* closed in L^∞. Similarly one can argue that the operator F is sequentially weakly*-weakly closed.

In the sequel we are going to employ approximation operators F_m as in [31]. Let Y_m be the space of linear splines on a uniform grid of $m+1$ points in $[0,1]$, which vanish at 0 and 1. By using the variational formulation, let $u_m(a) \in Y_m$ be the unique solution of

$$(a(u_m)_x, v_x)_{L^2} = (f, v)_{L^2} \text{ for all } v \in Y_m.$$

The operators F_m are defined as

$$F_m : \mathcal{D}(F) := \{a \in L^\infty[0,1] : a(x) \geq c > 0, \textit{ a.e.}\} \to L^2[0,1],$$

$$a \mapsto F_m(a) := u_m(a),$$

Then (6) holds for $\rho_m = m^{-2}$ cf. [12, Theorems 3.2.2, 3.2.5],

$$\|F_m(a) - F(a)\|_{L^2} = \|u_m(a) - u(a)\|_{L^2} = O(\|a\|_{L^\infty} \cdot m^{-2}).$$

Then by choosing the discretization of $L^\infty[0,1]$ as in Section 4, one obtains the convergence results of the discretized regularization method as in Example 4.13.

Acknowledgment

The authors thank Prof. J.B. Cooper (Kepler University, Linz) for providing them the counterexample in Section 4. They are grateful to S. Pereverzev (Radon Institute, Linz) for helpful discussions, and to A. Rieder (Karlsruhe University) and V. Vasin (Institute of Mathematics and Mechanics, Ekaterinburg) for interesting references. Also, they acknowledge the support by the Austrian Science Fund (FWF) within the national research networks Industrial Geometry, project 9203-N12, and Photoacoustic Imaging in Biology and Medicine, project S10505-N20 (C. P. and O. S.) and by an Elise Richter scholarship, project V82-N18 (E. R.).

References

[1] R. Adams. *Sobolev Spaces.* Academic Press, New York, 1975.

[2] R. Adams and J. Fournier. *Sobolev Spaces.* Academic Press, New York, 2003.

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

[4] R.S. Anderssen. The linear functional strategy for improperly posed problems. *Inverse problems, Oberwolfach*, pages 11–30, 1986.

[5] J. Borwein and A. Lewis. Convergence of best entropy estimates. *SIAM J. Optim.*, 1:191–205, 1991.

[6] M. Burger and S. Osher. Convergence rates of convex variational regularization. 20(5):1411–1421, 2004.

[7] M. Burger, E. Resmerita, and L. He. Error estimation for Bregman iterations and inverse scale space methods in image restoration. 81(2–3):109–135, 2007. Special Issue on Industrial Geometry.

[8] P. Bělík and M. Luskin. Approximation by piecewise constant functions in a BV metric. *Math. Models Methods Appl. Sci.*, 13(3):373–393, 2003. Dedicated to Jim Douglas, Jr. on the occasion of his 75th birthday.

[9] E. Casas, K. Kunisch, and C. Pola. Regularization by functions of bounded variation and applications to image enhancement. *Appl. Math. Optim.*, 40(2):229–257, 1999.

[10] A. Chambolle. An approximation result for special functions with bounded deformation. *Journal de Mathématiques Pures et Appliqués*, 83(7):929–954, 2004.

[11] G. Chavent. *Non Linear Least Squares For Inverse Problems. Scientific Computation.* Springer, 2009.

[12] P. Ciarlet. *The finite element method for elliptic problems.* Number 40 in SIAM Classics in Applied Mathematics. SIAM, Philadelphia, 2002.

[13] R.A. DeVore. Nonlinear approximation. *Acta Numerica*, 7:51–150, 1998.

[14] J. Duchi and Y. Singer. Efficient online and batch learning using forward backward splitting. *Journal of Machine Learning Research*, 10:28732908, 2009.

[15] H. Engl, A. Neubauer, and K. Kunisch. Convergence rates for tikhonov regularisation of non-linear ill-posed problems. *Inverse Problems*, 5(4):523, 1989.

[16] C. Fabre, J.-P. Puel and E. Zuazua. Contrôlabilité approchée de l'équation de la chaleur linéaire avec des contrôles de norme l^∞ minimale. (Approximate controllability for the linear heat equation with controls of minimal l^∞ norm) *C. R. Acad. Sci., Paris, Sér. I* 316:679–684, 1993.

[17] B.G. Fitzpatrick and S.L. Keeling. On approximation in total variation penalization for image reconstruction and inverse problems. 1997.

[18] K. Frick and O. Scherzer. Regularization of ill-posed linear equations by the non-stationary augmented lagrangian method. *J. Integral Equations Appl.*, to appear, 2009.

[19] M. Grasmair, M. Haltmeier, and O. Scherzer. Sparse regularization with l^q penalty term. *Inverse Problems*, 24(5):055020, 13, 2008.

[20] M. Gugat. Analytic solutions of L^∞ optimal control problems for the wave equation. *J. Optim. Theory Appl.* 114(2):397–421, 2002.

[21] M. Gugat, and G. Leugering. Regularization of L^∞-optimal control problems for distributed parameter systems. *Comput. Optim. Appl.* 22(2):151–192, 2002.

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

[23] R.B. Holmes. *Geometric Functional Analysis and Its Applications.* Springer, 1975.

[24] A. M. Holohan and G. Meinsma. On Inverse Problems in 2-norm and infinity-norm Controller Synthesis. *Proc. 1999 IEEE Conf. on Decision and Control*, San Diego, 1999.

[25] K. Ito and K. Kunisch. On the injectivity and linearization of the coefficient-to-solution mapping for elliptic boundary value problems. *J. Math. Anal. Appl.*, 188:1040–66, 1994.

[26] A.S. Leonov. Regularization of ill-posed problems in sobolev space w_1^1. *J. Inverse Ill-Posed Probl.*, 13:595—-619, 2005.

[27] A.S. Leonov. Total variation convergence of regularizing algorithms for solving ill-posed problems. *Zh. Vychisl. Mat. Mat. Fiz.*, 47:732–747, 2007.

[28] H. Matthies, G. Strang, and E. Christiansen. The saddle point of a differential program. In *Energy methods in finite element analysis*, pages 309–318. Wiley, Chichester, 1979.

[29] R.E. Megginson *An Introduction to Banach Space Theory.* Springer Verlag, New York, 1998.

[30] A. Neubauer. Tikhonov regularization for non-linear ill-posed problems: optimal convergence rates and finite-dimensional approximation. *Inverse Problems*, 5(4):541–557, 1989.

[31] A. Neubauer and O. Scherzer. Finite-dimensional approximation of tikhonov regularized solutions of nonlinear ill-posed problems. *Numer. Funct. Anal. Optim.*, (11):85–99, 1990.

[32] R. Plato and G. Vainikko. On the regularization of projection methods for solving ill-posed problems. *Numer. Math.*, 28:63–79, 1990.

[33] J. Qinian. Applications of the modified discrepancy principle to tikhonov regularization of nonlinear ill-posed problems. *SIAM J. Numer. Anal.*, 36:475–490, 1999.

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

[35] E. Resmerita and O. Scherzer. Error estimates for non-quadratic regularization and the relation to enhancement. *Inverse Problems*, 22(3):801–814, 2006.

[36] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging.* Applied Mathematical Sciences. Springer, 2008.

[37] P.-M. Suquet. Existence et régularité des solutions des équations de la plasticité. *C. R. Acad. Sci. Paris Sér. A-B*, 286(24):A1201–A1204, 1978.

[38] R. Temam. *Problèmes Mathématiques en Plasticité*, volume 12 of *Méthodes Mathématiques de l'Informatique [Mathematical Methods of Information Science].* Gauthier-Villars, Montrouge, 1983.

[39] R. Temam and G. Strang. Functions of bounded deformation. *Archive for Rational Mechanics and Analysis*, 75:7–21, 1980.

[40] A.N. Tikhonov and A.S. Leonov and A.G. Yagola. *Nonlinear Ill-Posed Problems* Champman & Hall, 1998.

[41] H. Zou and M. Yuan. The F-infinity-norm support vector machine. *Statistica Sinica*, 18:379-398, 2008.