Convergence rates for regularization functionals with polyconvex integrands

Clemens Kirisits[*], Otmar Scherzer[†]

July 28, 2017

Abstract

Convergence rates results for variational regularization methods typically assume the regularization functional to be convex. While this assumption is natural for scalar-valued functions, it can be unnecessarily strong for vector-valued ones. In this paper we focus on regularization functionals with polyconvex integrands. Even though such functionals are nonconvex in general, it is possible to derive linear convergence rates with respect to a generalized Bregman distance, an idea introduced by Grasmair in 2010. As a case example we consider the image registration problem.

1 Introduction

In this paper we consider solving ill-posed operator equations of the form

$$K(u) = v, \tag{1}$$

using Tikhonov-type regularization, which consists in approximation of a solution of (1) by the minimizer of the functional

$$\|K(u) - v\|^2 + \alpha \mathcal{R}(u). \tag{2}$$

Regularization theory is well-established when $\mathcal{R}$ is convex, and in particular when $\mathcal{R}(u) = \frac{1}{2}\|u - u_0\|^2$. See [9, 15, 23, 19, 20, 24, 25] for instance. Convergence rates results have been developed in [11, 12, 13, 16, 19, 20] among others. For nonconvex regularization functionals $\mathcal{R}$, however, only few results are available in the literature [3, 14, 26].

If the sought-for solution u is scalar-valued, then convexity of $\mathcal{R}$ is a natural condition, because it is closely linked to weak lower semicontinuity of $\mathcal{R}$. Yet if

[*]Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Linz, Austria

[†]Computational Science Center, University of Vienna, and Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Linz, Austria

$u : \Omega \subset \mathbb{R}^n \to \mathbb{R}^N$ is a *vector-valued* function, then properties strictly weaker than convexity are enough to ensure weak lower semicontinuity. On the other hand, using a nonconvex $\mathcal{R}$ raises the question of how to obtain convergence rates, since the most common approach involves Bregman distances, which in turn require $\mathcal{R}$ to be subdifferentiable. The aim of this article is to develop convergence rates results for regularization functionals with polyconvex integrands.

A function $f : \mathbb{R}^{N\times n} \to \mathbb{R}$ is polyconvex, if $f(A)$ can be written as a convex function of all subdeterminants of A. John Ball introduced this notion in the context of nonlinear elasticity, where convex stored energy functions are known to be too restrictive physically [2]. However, what lends importance to polyconvex functions even outside the field of elasticity is the fact that they render functionals of the form

$$\mathcal{R}(u) = \int_\Omega f(x, u(x), \nabla u(x))\, dx \tag{3}$$

weakly lower semicontinuous in $W^{1,p}(\Omega, \mathbb{R}^N)$.

Recently, the merits of polyconvex functions have been exploited in the field of image processing, in particular for image registration [4, 7, 17]. Practical applications of registration models are numerous, one of the most prominent being medical imaging [10, 22]. Registering two given images $I_1, I_2 : \Omega \to \mathbb{R}$ means finding a deformation $u : \Omega \to \mathbb{R}^n$ such that

$$I_1 \circ u = I_2. \tag{4}$$

The ill-posedness of this problem is typically overcome via variational regularization, that is, by minimizing a functional of the form

$$\mathcal{S}(I_1 \circ u, I_2) + \mathcal{R}(u),$$

where $\mathcal{S}$ measures the similarity between $I_1 \circ u$ and I_2. A regularization functional $\mathcal{R}$ with a polyconvex integrand can be a reasonable choice, if one models I_1 and I_2 as hyperelastic materials. However, in this case standard convergence rates results from regularization theory do not apply [19, 20]. The aim of this paper is to address this issue.

Outline. The next section (Sec. 2) introduces the most important concepts and fixes some notation. It consists of three parts. In the first part, Section 2.1, we introduce (generalized) Bregman distances. In Section 2.2, we review standard results on convergence rates for variational regularization of inverse problems in a Banach space setting. Section 2.3, briefly discusses polyconvex functions and their properties. Section 3 considers the image registration problem with polyconvex regularization from an inverse problems point of view. It also contains a specific example where in spite of nonconvex regularization the standard convergence rates result as stated in Sec. 2.2 applies. Finally, in Section 4 we define W_{poly}-Bregman distances for functionals with polyconvex integrands and state the corresponding convergence rates result.

2 Preliminiaries

2.1 Bregman distances

In this article U always denotes a Banach space with dual U^*. The dual pairing between $u \in U$ and $u^* \in U^*$ is denoted by $\langle u^*, u\rangle_{U^*,U}$. There are two notable special cases. If $U = U^* = \mathbb{R}^{N\times n}$, we write $u \cdot u^* = \sum_{i=1}^N \sum_{j=1}^n u^*_{ij} u_{ij}$. In the case of Lebesgue spaces (of possibly matrix-valued functions), we use dual brackets without subscripts and write

$$\langle u^*, u\rangle = \int u^*(x) \cdot u(x)\, dx.$$

Let $\Omega \subset \mathbb{R}^n$ be an open set, $1 \le p < \infty$ and let p^* be the Hölder conjugate of p. For a vector-valued Sobolev function $u \in U = W^{1,p}(\Omega, \mathbb{R}^N)$ we denote by ∇u the $N \times n$ matrix of weak partial derivatives of u. Every element u^* of U^* can be identified with a pair $(u^*_0, u^*_1) \in L^{p^*}(\Omega, \mathbb{R}^N \times \mathbb{R}^{N\times n})$ acting on $u \in U$ as

$$\langle u^*, u\rangle_{U^*,U} = \langle u^*_0, u\rangle + \langle u^*_1, \nabla u\rangle.$$

This is an immediate consequence of Thm. 3.9 in [1].

Let $\mathcal{R}$ be a function defined on U taking values in the extended reals $\mathbb{R} \cup \{\pm\infty\}$. Its *effective domain* $\operatorname{dom}\mathcal{R}$ is the set $\{u \in U : \mathcal{R}(u) < +\infty\}$. The *subdifferential* of $\mathcal{R}$ at $u \in U$ is defined as

$$\partial\mathcal{R}(u) = \begin{cases} \{u^* \in U^* : \mathcal{R}(v) \ge \mathcal{R}(u) + \langle u^*, v - u\rangle_{U^*,U} \text{ for all } v \in U\}, & \mathcal{R}(u) \in \mathbb{R} \\ \emptyset, & \mathcal{R}(u) \notin \mathbb{R}. \end{cases}$$

Note that we have not assumed $\mathcal{R}$ to be convex. If $\partial\mathcal{R}(u) \ne \emptyset$, then $\mathcal{R}$ is said to be *subdifferentiable* at u and elements $u^* \in \partial\mathcal{R}(u)$ are called *subgradients*. Recall Fermat's rule: A proper function $\mathcal{R}$ attains its minimum at $u \in U$, if and only if $0 \in \partial\mathcal{R}(u)$. Let $u \in \operatorname{dom}\mathcal{R}$ and $u^* \in \partial\mathcal{R}(u)$. The *Bregman distance* associated to $\mathcal{R}$ at (u, u^*) is defined as

$$D_{u^*}(v; u) = \mathcal{R}(v) - \mathcal{R}(u) - \langle u^*, v - u\rangle_{U^*,U}.$$

The following lemma justifies the use of the Bregman distance as a similarity measure.

Lemma 2.1. *The Bregman distance is nonnegative and satisfies* $D_{u^*}(u; u) = 0$.

The Bregman distance is only defined at points where $\mathcal{R}$ has a subgradient. For convex functions these points can be characterized easily. The first two of the following three lemmas are classical results on subdifferentiability of convex functions. The third one deals with the special case of integral functionals on Sobolev spaces.

Lemma 2.2. *Let* $\mathcal{R} : U \to \mathbb{R} \cup \{\pm\infty\}$ *be a convex function. If* $\mathcal{R}$ *is finite and continuous at one point* $\bar{u} \in U$*, then* $\partial\mathcal{R}(u) \ne \emptyset$ *for all* $u \in \operatorname{int}\operatorname{dom}\mathcal{R}$.

Proof. See Proposition 5.2 in Chapter I of [8]. □

Lemma 2.3. *If $\mathcal{R} : U \to \mathbb{R} \cup \{\pm\infty\}$ is proper, convex and lower semicontinuous, then the set $\{u \in U : \partial\mathcal{R}(u) \neq \emptyset\}$ is dense in $\operatorname{dom}\mathcal{R}$.*

Proof. See Corollary 6.2 in Chapter I of [8]. □

Lemma 2.4. *Let $\Omega \subset \mathbb{R}^n$ be an open set and let*

$$f : \Omega \times \mathbb{R}^N \times \mathbb{R}^{N\times n} \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$$

be a Carathéodory function. Assume that, for almost every $x \in \Omega$, the map $(u, A) \mapsto f(x, u, A)$ is convex and differentiable throughout its effective domain. Let $p \in [1, \infty)$ and define the following functional on $W^{1,p}(\Omega, \mathbb{R}^N)$

$$\mathcal{R}(v) = \int_\Omega f(x, v(x), \nabla v(x))\, dx.$$

Denote by $\nabla_{u,A} f$ the gradient of f with respect to its second and third variables. If $v \in \operatorname{dom}\mathcal{R}$ and the function

$$x \mapsto \nabla_{u,A} f(x, v(x), \nabla v(x))$$

lies in $L^{p^*}(\Omega, \mathbb{R}^N \times \mathbb{R}^{N\times n})$, then this function is a subgradient of $\mathcal{R}$ at v.

Proof. This is a direct consequence of Lemma 4.1 in Chapter X of [8]. □

Bregman distances play an important role for convex variational regularization, as they can be used to measure the rate of convergence of regularized solutions. In this paper we want to consider *nonconvex* regularization functionals. Nonconvex functions, however, are not subdifferentiable in general and therefore the associated Bregman distance is not of much use. One way to overcome this problem was explored by Grasmair in [14]. It is based on an abstract version of convexity theory where, essentially, the dual space U^* is replaced by some other set W of (extended) real-valued functions on U. See also Chapter 8 of [21].

Definition 2.1. Let W be a family of real-valued functions defined on U. Following [14, 21] we define the *W-subdifferential* of $\mathcal{R}$ at $u \in U$ as

$$\partial_W \mathcal{R}(u) = \begin{cases} \{w \in W : \mathcal{R}(v) \geq \mathcal{R}(u) + w(v) - w(u) \text{ for all } v \in U\}, & \mathcal{R}(u) \in \mathbb{R} \\ \emptyset, & \mathcal{R}(u) \notin \mathbb{R}. \end{cases}$$

For $w \in \partial_W \mathcal{R}(u)$ the corresponding *W-Bregman distance* is given by

$$D_w^W(v; u) = \mathcal{R}(v) - \mathcal{R}(u) - w(v) + w(u). \tag{5}$$

Clearly, the U^*-subdifferential and the U^*-Bregman distance coincide with their classical counterparts.

Lemma 2.5. *The W-Bregman distance is nonnegative and satisfies*

$$D_w^W(u; u) = 0.$$

When using a W-Bregman distance to measure convergence rates, it is important to be able to characterize its domain of definition, that is, the set $\{u \in U : \partial_W \mathcal{R}(u) \neq \emptyset\}$. For the classical Bregman distance this characterization is done by Lemmas 2.2, 2.3 and 2.4. Below, in Section 4, we introduce a particular instance of a set W. This set, denoted by W_{poly}, is designed specifically for functionals with polyconvex integrands. In Lemma 4.1, we address the problem of characterizing the domain of definition of the W_{poly}-Bregman distance by proving a result analogous to 2.4.

2.2 Variational regularization on Banach spaces

Let U, V be Banach spaces and $K : \mathcal{D}(K) \subset U \to V$. We consider the inverse problem of finding $u \in U$ such that

$$K(u) = v^\dagger. \tag{6}$$

Exact data $v^\dagger \in \operatorname{ran} K \subset V$ are assumed to be available as noisy measurements $v^\delta \in V$ only, satisfying $\|v^\dagger - v^\delta\| \leq \delta$ for some $\delta \geq 0$. Since such problems are ill-posed in general, regularization is needed for the approximate inversion of K. Variational regularization consists in minimization of a functional of the form

$$u \mapsto \mathcal{T}_\alpha(u; v^\delta) = \|K(u) - v^\delta\|^q + \alpha \mathcal{R}(u), \tag{7}$$

where $q \geq 1$, $\alpha > 0$, $\|\cdot\|$ denotes the norm on V and $\mathcal{R} : U \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$ is such that $\operatorname{dom} \mathcal{R} \cap \mathcal{D}(K) \neq \emptyset$. We set $\mathcal{T}_\alpha(u; v^\delta) = +\infty$, if $u \notin \mathcal{D}(K)$. We call this variational approach a *well-defined regularization method*, if it fulfils the following requirements.

Existence: $\mathcal{T}_\alpha(\cdot; v^\delta)$ has a minimizer u_α^δ for every $v^\delta \in V$ and $\alpha > 0$.

Stability: The inversion $v^\delta \mapsto u_\alpha^\delta$ is continuous.

Convergence: There exists a parameter choice rule $\alpha : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ such that regularized solutions u_α^δ converge to a solution of (6) as $\delta \to 0$.

The last point in particular requires that the set of *exact solutions* $K^{-1}(v^\dagger)$ be nonempty. Exact solutions which minimize $\mathcal{R}$, that is, elements of the set

$$\arg\min\{\mathcal{R}(u) : u \in K^{-1}(v^\dagger)\},$$

are called *$\mathcal{R}$-minimizing solutions*. The following theorem gives conditions for when minimization of (7) is a well-defined regularization method (cf. Section 3.2 of [19]).

Theorem 2.1. *Endow the Banach spaces U and V with topologies weaker than the respective norm topologies. Assume that the following four statements hold with respect to these topologies:*

1. *The sublevel sets of $\mathcal{T}_\alpha(\cdot; v^\dagger)$ are sequentially precompact.*

2. *$\|\cdot\|$ is sequentially lower semicontinuous.*

3. *$\mathcal{R}$ is sequentially lower semicontinuous.*

4. *The sublevel sets of $\mathcal{T}_\alpha(\cdot; v^\dagger)$ are sequentially closed and K is sequentially continuous there.*

Then the functional $\mathcal{T}_\alpha(\cdot; v^\delta)$ has a minimum for all $\alpha > 0$ and $v^\delta \in V$. Moreover minimization of $\mathcal{T}_\alpha$ is continuous in the following sense. Whenever $\|v_k - v^\delta\| \to 0$, then every sequence (u_k), $u_k \in \arg\min \mathcal{T}_\alpha(\cdot; v_k)$, has a converging subsequence and the limit of every such sequence is a minimizer of $\mathcal{T}_\alpha(\cdot; v^\delta)$. Assume, in addition, that

5. *there is an exact solution in $\operatorname{dom} \mathcal{R}$ and*

6. *the parameter choice rule $\alpha : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ satisfies $\alpha(\delta) \to 0$ and $\delta^q/\alpha(\delta) \to 0$ as $\delta \to 0$.*

Then, whenever $\delta_k \to 0$, every sequence (u_k), $u_k \in \arg\min \mathcal{T}_\alpha(\cdot; v^{\delta_k})$, has a converging subsequence and the limit of every such sequence is an $\mathcal{R}$-minimizing solution.

Remark 2.1. Note that convexity of $\mathcal{R}$ is not required for a well-defined regularization method.

In principle convergence of regularized solutions can be arbitrarily slow. Therefore it is useful to have a bound in terms of δ on the discrepancy between regularized and exact solution. In a Banach space setting a typical discrepancy measure is the Bregman distance associated to the regularization functional [5]. Concerning convergence rates for variational regularization in Banach spaces we have the following result (cf. [16] or Section 3.2 of [19]).

Theorem 2.2. *Let the assumptions of Theorem 2.1 hold. In addition assume that $\mathcal{R}$ has a subgradient u^* at an $\mathcal{R}$-minimizing solution $u^\dagger$ and that there are constants $\beta_1 \in [0, 1)$, $\beta_2, \bar{\alpha} > 0$ and $\rho > \bar{\alpha}\mathcal{R}(u^\dagger)$ such that*

$$\langle u^*, u^\dagger - u \rangle \leq \beta_1 D_{u^*}(u; u^\dagger) + \beta_2 \|K(u) - v^\dagger\| \tag{8}$$

holds for all u with $\mathcal{T}_{\bar{\alpha}}(u; v^\dagger) \leq \rho$.

If $q > 1$, assume $\alpha(\delta) \sim \delta^{q-1}$. Then

$$D_{u^*}(u_\alpha^\delta; u^\dagger) = O(\delta) \quad \text{and} \quad \|K(u_\alpha^\delta) - v^\delta\| = O(\delta).$$

If $q = 1$, assume $\alpha(\delta) \sim \delta^\epsilon$ for $\epsilon \in (0, 1)$. Then

$$D_{u^*}(u_\alpha^\delta; u^\dagger) = O(\delta^{1-\epsilon}) \quad \text{and} \quad \|K(u_\alpha^\delta) - v^\delta\| = O(\delta).$$

Remark 2.2. In contrast to [16, 19] we have not assumed $\mathcal{R}$ to be convex and have not added any other assumption in its place. This does not change the theorem's validity. However, for a general nonconvex $\mathcal{R}$ the condition $u^* \in \partial\mathcal{R}(u^\dagger)$ cannot be expected to hold except in certain special cases. (Recall that standard results on subdifferentiability, such as Lemmas 2.2, 2.3 and 2.4, require $\mathcal{R}$ to be convex.) Example 3.2 below is devoted to such a special case. That is, we construct a well-defined regularization method where $\mathcal{R}$ is not convex but condition (8) is still satisfied.

Remark 2.3. Moreover, Example 3.2 exploits the following observation. If the regularization functional is chosen "perfectly", that is, it has a global minimizer that is also an exact solution, then condition (8) is always satisfied: Assume that $\bar{u}^\dagger$ is such a solution, that is, $K(\bar{u}^\dagger) = v^\dagger$ and $0 \in \partial\mathcal{R}(\bar{u}^\dagger)$. Then, with $u^* = 0$ inequality (8) becomes

$$0 \leq \beta_1(\mathcal{R}(u) - \mathcal{R}(\bar{u}^\dagger)) + \beta_2\|K(u) - v^\dagger\|,$$

which is true for all $u \in U$ and all nonnegative β_1, β_2.

2.3 Polyconvex functions

When considering regularization functionals of the form

$$\mathcal{R}(u) = \int_\Omega f(x, u(x), \nabla u(x))\, dx$$

on $U = W^{1,p}(\Omega, \mathbb{R}^N)$ one major concern is how to ensure weak lower semicontinuity. While convexity of f in its third argument is essentially an equivalent condition in the scalar setting ($n = 1$ or $N = 1$), convexity is unnecessarily restrictive when $n > 1$ and $N > 1$. There, the weaker notion of quasiconvexity is already sufficient. Unfortunately, quasiconvexity can be a difficult property to verify. Polyconvexity, however, while still weaker than convexity, is stronger than quasiconvexity and easier to work with.

Let $N, n \in \mathbb{N}$ and let $N \wedge n = \min(N, n)$. For $A \in \mathbb{R}^{N\times n}$ and $1 \leq s \leq N \wedge n$ denote by $\operatorname{adj}_s(A)$ the matrix consisting of all $s \times s$ minors of A. Note that $\operatorname{adj}_1(A) = A$ and $\operatorname{adj}_s(A) \in \mathbb{R}^{\sigma(s)}$, where $\sigma(s) = \binom{N}{s}\binom{n}{s}$. Set $\tau(N, n) = \sum_{s=1}^{N\wedge n} \sigma(s)$ and denote by $T : \mathbb{R}^{N\times n} \to \mathbb{R}^{\tau(N,n)}$ the function that maps a matrix to the vector containing all its minors, which with a slight abuse of notation can be written as

$$T(A) = (A, \operatorname{adj}_2(A), \ldots, \operatorname{adj}_{N\wedge n}(A)).$$

A function $f : \mathbb{R}^{N\times n} \to \mathbb{R} \cup \{+\infty\}$ is called *polyconvex*, if there exists a convex function $F : \mathbb{R}^{\tau(N,n)} \to \mathbb{R} \cup \{+\infty\}$ satisfying $f = F \circ T$. Notice that this F is not unique in general. Clearly, every convex function is polyconvex. If $N = 1$ or $n = 1$, then also the converse holds. If $n = N > 1$, an example of a polyconvex function which is not convex is $f(A) = |\det A|^2$. See Chapter 5 in [6] for more details on polyconvex functions.

The following result is a special case of the more general Theorem 8.16 in [6].

Lemma 2.6. *Let $\Omega \subset \mathbb{R}^n$ be a bounded open set with Lipschitz boundary and let*

$$F : \Omega \times \mathbb{R}^N \times \mathbb{R}^{\tau(N,n)} \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$$

be a Carathéodory function such that the map $A \mapsto F(x, u, A)$ is convex for almost every $x \in \Omega$ and every $u \in \mathbb{R}^N$. Then, for $p > N \wedge n$, the functional

$$u \mapsto \int_\Omega F(x, u(x), T(\nabla u(x)))\, dx$$

is sequentially weakly lower semicontinuous in $W^{1,p}(\Omega, \mathbb{R}^N)$.

In the last part of this paper we will make use of the following variant of the map T. Set $\tau_2(N,n) = \sum_{s=2}^{N\wedge n} \sigma(s)$. We denote by $T_2 : \mathbb{R}^{N\times n} \to \mathbb{R}^{\tau_2(N,n)}$ the function defined by

$$T_2(A) = (\mathrm{adj}_2(A), \ldots, \mathrm{adj}_{N\wedge n}(A)).$$

3 Image registration

In this section we treat the image registration problem from an inverse problems perspective. First, by applying Theorem 2.1 we show that minimization of

$$\|I_2 \circ u - I_1\|^q + \alpha \mathcal{R}(u),$$

where $\mathcal{R}$ is a first order functional with polyconvex integrand, constitutes a well-defined regularization method. Second, we highlight a particular situation where, in spite of $\mathcal{R}$ being nonconvex, Theorem 2.2 applies as well.

Let $\Omega \subset \mathbb{R}^n$ be a bounded open set with Lipschitz boundary. Given a target image $I_1 : \Omega \to \mathbb{R}$ and a reference image $I_2 : \Omega \to \mathbb{R}$ the model equation for the image registration problem reads

$$I_2 \circ u = I_1,$$

where $u : \Omega \to \mathbb{R}^n$ is an unknown deformation of the image domain. We interpret this as a particular instance of the abstract operator equation (6). Thus K is the composition operator that sends every deformation u to the deformed reference image $I_2 \circ u = K(u)$. Note that in Section 2.2 we have implicitly assumed that the operator be known exactly. Therefore, I_2 is known exactly, whereas the exact target image $I_1^\dagger$, i.e. the exact data, is available only as noisy measurements I_1^δ.

Theorem 3.1. *Let $p > n$ and $q \geq 1$. Endow $U = W^{1,p}(\Omega, \mathbb{R}^n)$, with its weak and $V = L^q(\Omega)$ with its strong topology. Assume $I_2 \in C^0(\bar{\Omega})$ and define the operator*

$$K : \mathcal{D}(K) \subset U \to V, \quad u \mapsto K(u) = I_2 \circ u$$

with domain $\mathcal{D}(K) = \{u \in U : u(\Omega) \subset \bar{\Omega}\}$. Let

$$F : \Omega \times \mathbb{R}^n \times \mathbb{R}^{\tau(n,n)} \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$$

be a Carathéodory function such that, for almost every $x \in \Omega$ and every $u \in \mathbb{R}^n$, the map $\xi \mapsto F(x,u,\xi)$ is convex and

$$F(x,u,T(A)) \geq c|A|^p \tag{9}$$

holds for every $A \in \mathbb{R}^{n\times n}$ and some $c > 0$. For $u \in U$ define

$$\mathcal{R}(u) = \int_\Omega F(x, u(x), T(\nabla u(x)))\, dx$$

and assume that $\operatorname{dom}\mathcal{R} \cap \mathcal{D}(K)$ is not empty. Then, minimization of

$$\mathcal{T}_\alpha(u; I_1^\delta) = \|K(u) - I_1^\delta\|^q_{L^q(\Omega)} + \alpha\mathcal{R}(u), \quad \alpha > 0, \tag{10}$$

is a well-defined regularization method in the sense of Theorem 2.1.

Proof. We show that all the assumptions of Theorem 2.1 are satisfied.

Item 2: The norm is continuous and therefore also semicontinuous.

Item 3: Lower semicontinuity of $\mathcal{R}$ follows from Lemma 2.6.

Item 1: We have to show that the sublevel sets of $\mathcal{T}_\alpha(\cdot; I_1)$ are weakly sequentially precompact. Let $\alpha, M > 0$, $I_1 \in V$ and $(u_k) \subset U$ with $\mathcal{T}_\alpha(u_k; I_1) \leq M$ for $k \geq 1$. Then, in particular, $(u_k) \subset \mathcal{D}(K)$ and therefore $u_k(\Omega) \subset \bar{\Omega}$ for all k. Since Ω is bounded, the sequence (u_k) is bounded in $L^p(\Omega, \mathbb{R}^n)$. The lower bound (9) on F yields boundedness of (u_k) in U. Since $p > 1$, U is reflexive and therefore, by the Eberlein-Šmulian theorem, every bounded sequence has a weakly convergent subsequence. Thus, the sublevel sets of $\mathcal{T}_\alpha(\cdot; I_1)$ are weakly sequentially precompact.

Item 4: We need to verify that the sublevel sets of $\mathcal{T}_\alpha(\cdot; I_1)$ are weakly sequentially closed and that K is weak-strong sequentially continuous on these sets. Let again $\mathcal{T}_\alpha(u_k; I_1) \leq M$ for $k \geq 1$ and assume $u_k \rightharpoonup \bar{u}$ for some $\bar{u} \in U$. The compact embedding of U into $C^0(\bar{\Omega}, \mathbb{R}^n)$ implies that $u_k \to \bar{u}$ uniformly and $\bar{u} \in \mathcal{D}(K)$. Since $I_2 \in C^0(\bar{\Omega})$, the sequence $(I_2 \circ u_k)$ converges uniformly to $I_2 \circ \bar{u}$ and, because Ω is bounded, it also converges in $L^q(\Omega)$. Finally, continuity of $\|\cdot\|^q_{L^q}$ and weak lower semicontinuity of $\mathcal{R}$ gives

$$\mathcal{T}_\alpha(\bar{u}; I_1) \leq \liminf_{k\to\infty} \mathcal{T}_\alpha(u_k; I_1) \leq M.$$

Thus we have shown that the sublevel sets of $\mathcal{T}_\alpha(\cdot; I_1)$ are weakly sequentially closed and that K is weak-strong sequentially continuous there. □

Example 3.1. Let $n = 3$. For $\gamma_1, \gamma_2 \geq 0$ consider the following regularization term

$$\mathcal{R}(u) = E_{\text{vol}}(u) + \gamma_1 E_{\text{mem}}(u) + \gamma_2 E_{\text{bend}}(u),$$

where E_{vol}, E_{mem} and E_{bend} are the volume, membrane and bending energies, respectively, from [17]. They all have polyconvex integrands and, in addition, the volume energy satisfies the coercivity estimate (9). Moreover, the identity transformation lies in $\mathcal{D}(K) \cap \operatorname{dom} \mathcal{R}$. It now follows from Theorem 3.1 that minimization of

$$\|I_2 \circ u - I_1^\delta\|_{L^2(\Omega)}^2 + \alpha \mathcal{R}(u)$$

over $U = W^{1,p}(\Omega, \mathbb{R}^3)$, $p > 3$, is a well-defined regularization method. Note, however, that in [17] boundary conditions are imposed and the data matching term is weighted with a nonnegative cutoff function whose support is a proper subset of Ω.

As already pointed out in Remark 2.2, while Theorem 2.2 is in principle applicable to regularization methods like (10), in most cases, due to the nonconvexity of $\mathcal{R}$, it is unlikely to actually apply. Below, we use the idea from Remark 2.3 to construct an instance of a registration problem where Theorem 2.2 does apply however.

Example 3.2. Let $n = 2$. Assume that I_2 is a rotated version of the exact data $I_1^\dagger$. That is, there is a deformation $u_R \in U$, given by $u_R(x) = Rx$ for some $R \in SO(2)$, such that $\|I_2 \circ u_R - I_1^\dagger\|_{L^q(\Omega)} = 0$. Of course, u_R must lie in $\mathcal{D}(K)$, which in this case translates to Ω being invariant with respect to the rotation R. Below we construct a nonconvex regularization functional $\mathcal{R}$ which not only satisfies all requirements from Theorem 3.1 but which is also minimal for rotations. It then follows that $0 \in \partial \mathcal{R}(u_R)$ and, by Remark 2.3, Theorem 2.2 applies.

For $u \in W^{1,p}(\Omega, \mathbb{R}^2)$, $p > 2$, we define

$$\mathcal{R}(u) = \int_\Omega f(\nabla u(x))\, dx,$$

where

$$f(A) = \operatorname{tr}\Big[(A^\top A)^{p/2}\Big] + p e^{1 - \det A}$$

for all $A \in \mathbb{R}^{2\times 2}$. This particular choice of integrand is loosely inspired by the tangential distortion energy from [17]. Next we verify the requirements of Theorem 3.1.

The identity deformation lies in the set $\operatorname{dom} \mathcal{R} \cap \mathcal{D}(K)$. Hence it is nonempty. The coercivity estimate (9) follows from

$$f(A) \geq \operatorname{tr}\Big[(A^\top A)^{p/2}\Big] = \lambda_1^p + \lambda_2^p \geq c(\lambda_1^2 + \lambda_2^2)^{p/2} = c|A|^p.$$

Here λ_1, λ_2 are the singular values of A and $c > 0$ is a constant whose existence is guaranteed by the equivalence of norms in finite dimensions. Convexity of the maps $x \mapsto e^{1-x}$ and $A \mapsto \lambda_1^p + \lambda_2^p$ (cf. [18, Lemma 3.11]) yields polyconvexity of f. Therefore, minimization of (10) with $\mathcal{R}$ as specified above is a well-defined regularization method.

To verify minimality on $SO(2)$ it is convenient to rewrite f in terms of its signed singular values $\mu_1 = \operatorname{sgn}(\det A)\lambda_1$, $\mu_2 = \lambda_2$:

$$f(A) = |\mu_1|^p + \mu_2^p + p e^{1 - \mu_1 \mu_2}.$$

Now minimality on $SO(2)$ translates to minimality for $\mu_1 = \mu_2 = 1$, which is easy to check. Thus, R minimizes f and consequently u_R minimizes $\mathcal{R}$. Since u_R is also an exact solution, the convergence rates result Thm. 2.2 applies.

The fact that f only depends on signed singular values is actually equivalent to f being $SO(2) \times SO(2)$ invariant, which is a desirable property in itself. See [6, Sec. 5.3.3] for more details.

4 Generalized Bregman distances for functionals with polyconvex integrands

The previous example has shown that there are problems where even nonconvex regularization can lead to a linear convergence rate in the classical Bregman distance. In general, however, mainly due to a lack of subdifferentiability, this cannot be expected. Therefore, we wish to find a weaker notion of subdifferentiability that is better suited for functionals with polyconvex integrands. In doing so we follow Grasmair's approach from [14], which for our purposes boils down to finding a set W that takes the place of the dual U^* in the definition of the subdifferential. This set is chosen such that we can prove a W-subdifferentiability result, similar to Lemma 2.4, for a certain class of functionals with polyconvex integrands. In further consequence, the associated W-Bregman distance allows us to translate the classical convergence rates result Theorem 2.2 to the polyconvex setting.

Definition 4.1. Let $\Omega \subset \mathbb{R}^n$ be open. For $p \geq N \wedge n$ set $U = W^{1,p}(\Omega, \mathbb{R}^N)$. Recall the notation from Section 2.3 and observe that for $u \in U$ by Hölder's inequality we have

$$T_2(\nabla u) \in \prod_{s=2}^{N \wedge n} L^{\frac{p}{s}}(\Omega, \mathbb{R}^{\sigma(s)}) =: S_2. \tag{11}$$

Therefore, we let W_{poly} be the set of all functions $w : U \to \mathbb{R}$ for which there is a pair $(u^*, v^*) \in U^* \times S_2^*$ such that

$$w(u) = \langle u^*, u \rangle_{U^*,U} + \langle v^*, T_2(\nabla u) \rangle_{S_2^*, S_2}$$

for all $u \in U$.

Remark 4.1. The previous definition can be regarded as a natural one in the following sense. The dual U^* basically consists of functions that act on all pairs $(u, \nabla u) \in L^p(\Omega, \mathbb{R}^N) \times L^p(\Omega, \mathbb{R}^{N \times n})$ in a linear and bounded way. Similarly, the set W_{poly} consists of functions acting on $(u, T(\nabla u))$ in a linear and bounded way.

Remark 4.2. Identifying $u^* \in U^*$ with $(u^*, 0) \in W_{\mathrm{poly}}$ we can regard W_{poly} as a superset of U^*. Hence the generalized subdifferential

$$\partial_{\mathrm{poly}} \mathcal{R}(u) = \begin{cases} \{w \in W_{\mathrm{poly}} : \mathcal{R}(v) \geq \mathcal{R}(u) + w(v) - w(u) \text{ for all } v \in U\}, & \mathcal{R}(u) \in \mathbb{R} \\ \emptyset, & \mathcal{R}(u) \notin \mathbb{R}. \end{cases}$$

can be regarded as a superset of the classical one $\partial\mathcal{R}(u)$ for all functionals $\mathcal{R} : U \to \mathbb{R} \cup \{\pm\infty\}$ and $u \in U$. Put differently, every functional which is subdifferentiable in the classical sense is also W_{poly}-subdifferentiable. This is analogous to polyconvexity being a weaker notion than convexity.

Remark 4.3. Let $\bar{u} \in \operatorname{dom}\mathcal{R}$, $u \in U$ and $w = (u^*, v^*) \in \partial_{\mathrm{poly}}\mathcal{R}(\bar{u})$. The associated W_{poly}-Bregman distance can be written as

$$\begin{aligned} D_w^{\mathrm{poly}}(u;\bar{u}) &= \mathcal{R}(u) - \mathcal{R}(\bar{u}) - w(u) + w(\bar{u}) \\ &= \mathcal{R}(u) - \mathcal{R}(\bar{u}) - \langle u^*, u - \bar{u}\rangle_{U^*,U} - \langle v^*, T_2(\nabla u) - T_2(\nabla \bar{u})\rangle_{S_2^*,S_2}. \end{aligned}$$

Note that the first three terms in the second line correspond to the classical Bregman distance at $(\bar{u}, u^*)$, but their sum can be negative now, since $u^* \notin \partial\mathcal{R}(\bar{u})$ in general. If however $v^* = 0$, then $u^* \in \partial\mathcal{R}(\bar{u})$ and the classical and W-Bregman distances coincide. That is,

$$D_w^{\mathrm{poly}}(u;\bar{u}) = D_{u^*}(u;\bar{u})$$

for all $u \in U$.

The following statement justifies our definition of W_{poly} and is analogous to Lemma 2.4.

Lemma 4.1. *Let $\Omega \subset \mathbb{R}^n$ be an open set and let*

$$F : \Omega \times \mathbb{R}^N \times \mathbb{R}^{\tau(N,n)} \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$$

be a Carathéodory function. Assume that, for almost every $x \in \Omega$, the map $(u,\xi) \mapsto F(x,u,\xi)$ is convex and differentiable throughout its effective domain. Let $p \in [1,\infty)$ and define the following functional on $U = W^{1,p}(\Omega,\mathbb{R}^N)$

$$\mathcal{R}(u) = \int_\Omega F(x, u(x), T(\nabla u(x)))\, dx.$$

If $\mathcal{R}(\bar{v}) \in \mathbb{R}$ and the function $x \mapsto \nabla_{u,\xi} F(x, \bar{v}(x), T(\nabla \bar{v}(x)))$ lies in

$$L^{p^*}(\Omega,\mathbb{R}^N) \times \prod_{s=1}^{N\wedge n} L^{(\frac{p}{s})^*}(\Omega,\mathbb{R}^{\sigma(s)}), \tag{12}$$

then this function is a W_{poly}-subgradient of $\mathcal{R}$ at $\bar{v}$.

Proof. Since, for almost every $x \in \Omega$, the map $(u,\xi) \mapsto F(x,u,\xi)$ is convex and differentiable throughout its effective domain, it is subdifferentiable. Therefore for every $(v,\zeta) \in \operatorname{dom} F(x,\cdot,\cdot)$ we have

$$\begin{aligned} F(x,w,\eta) &\geq F(x,v,\zeta) + \nabla_{u,\xi} F(x,v,\zeta) \cdot (w - v, \eta - \zeta) \\ &= F(x,v,\zeta) + \nabla_u F(x,v,\zeta) \cdot (w - v) + \nabla_\xi F(x,v,\zeta) \cdot (\eta - \zeta) \end{aligned}$$

for all $(w,\eta) \in \mathbb{R}^N \times \mathbb{R}^{\tau(N,n)}$. In particular, for functions $\bar{v}, v \in U$, $\mathcal{R}(\bar{v}) \in \mathbb{R}$, we get

$$\begin{aligned} F(x, v(x), T(\nabla v(x))) \geq F(x, \bar{v}(x), T(\nabla \bar{v}(x))) \\ &+ \nabla_u F(x, \bar{v}(x), T(\nabla \bar{v}(x))) \cdot (v(x) - \bar{v}(x)) \\ &+ \nabla_\xi F(x, \bar{v}(x), T(\nabla \bar{v}(x))) \cdot (T(\nabla v(x)) - T(\nabla \bar{v}(x))) \end{aligned}$$

for almost every $x \in \Omega$. Integration over Ω gives

$$\begin{aligned} \mathcal{R}(v) \geq \mathcal{R}(\bar{v}) + \int_\Omega \big[&\nabla_u F(x, \bar{v}(x), T(\nabla \bar{v}(x))) \cdot (v(x) - \bar{v}(x)) \\ &+ \nabla_\xi F(x, \bar{v}(x), T(\nabla \bar{v}(x))) \cdot (T(\nabla v(x)) - T(\nabla \bar{v}(x))) \big] \, dx. \end{aligned}$$

Considering that $v - \bar{v} \in L^p(\Omega, \mathbb{R}^N)$ and

$$T(\nabla v) - T(\nabla \bar{v}) \in \prod_{s=1}^{N \wedge n} L^{\frac{p}{s}}(\Omega, \mathbb{R}^{\sigma(s)}),$$

the integral on the right hand side is well-defined, if

$$x \mapsto \nabla_u F(x, \bar{v}(x), T(\nabla \bar{v}(x))) \in L^{p^*}(\Omega, \mathbb{R}^N)$$

and

$$x \mapsto \nabla_\xi F(x, \bar{v}(x), T(\nabla \bar{v}(x))) \in \prod_{s=1}^{N \wedge n} L^{(\frac{p}{s})^*}(\Omega, \mathbb{R}^{\sigma(s)}),$$

which is just what we have assumed in (12). $\square$

Example 4.1. Let $N = n = 2$. Then $T(A) = (A, \det A)$ for $A \in \mathbb{R}^{2\times 2}$. Define an integrand by $F(x, u, A, \det A) = F(\det A) = (\det A)^2$. If $p = 4$, then for all $u \in U = W^{1,4}(\Omega, \mathbb{R}^2)$ we have

$$\mathcal{R}(u) = \int_\Omega (\det \nabla u(x))^2 \, dx \in \mathbb{R}$$

and the function $x \mapsto F'(\det \nabla u(x)) = 2 \det \nabla u(x)$ lies in $L^2(\Omega)$. By Lemma 4.1 functional $\mathcal{R}$ is W_{poly}-subdifferentiable everywhere.

Example 4.2. Let $p > N = n \geq 2$, $q > 1$, and let $\Omega \subset \mathbb{R}^n$ be bounded. Consider the integrand $F(x, u, T(A)) = F(A, \det A) = |A|^p/p + |\det A|^q/q$. If $\bar{v} \in W^{1,\infty}(\Omega, \mathbb{R}^n)$, then clearly $\mathcal{R}(\bar{v}) \in \mathbb{R}$. In addition

$$x \mapsto \nabla_\xi F(\nabla \bar{v}(x), \det \nabla \bar{v}(x)) = (|\nabla \bar{v}(x)|^{p-2} \nabla \bar{v}(x), |\det \nabla \bar{v}(x)|^{q-2} \det \nabla \bar{v}(x))$$

lies in L^∞. Therefore, $\mathcal{R}$ has a W_{poly}-subgradient everywhere on $W^{1,\infty}(\Omega, \mathbb{R}^n) \subset U = W^{1,p}(\Omega, \mathbb{R}^n)$, which implies that the associated W_{poly}-Bregman distance is defined on a dense subset of U. In addition, the functional satisfies the coercivity estimate (9) and is weakly lower semicontinuous in $W^{1,p}$ according to Lemma 2.6. Thus, at least from a theoretical perspective, the functional $\mathcal{R}$ is well-suited for regularizing inverse problems with $\mathbb{R}^n$-valued unknowns. Since $\mathcal{R}$ is not convex, however, it is not covered by most of existing regularization theory [19, 20].

Remark 4.4. If a functional $\mathcal{R}$ on U is W_{poly}-subdifferentiable at $u \in U$, then it is also locally W_{poly}-convex at u in the sense of [14]. This means that $\mathcal{R}(u) = \mathcal{R}^{**}(u)$, where the asterisk indicates generalized Fenchel conjugation with respect to W_{poly}. (See Section 2 in [14] or [21] for more details.) Therefore, Lemma 4.1 also provides sufficient conditions for a functional to be locally W_{poly}-convex.

The next theorem shows that standard convergence rates results can be carried over to the W_{poly}-Bregman distance.

Theorem 4.1. *Let the assumptions of Theorem 2.1 hold with* $U = W^{1,p}(\Omega, \mathbb{R}^N)$*. In addition assume that* $\mathcal{R}$ *has a* W_{poly}*-subgradient* w *at an* $\mathcal{R}$*-minimizing solution* $u^\dagger$ *and that there are constants* $\beta_1 \in [0,1)$, $\beta_2, \bar{\alpha} > 0$ *and* $\rho > \bar{\alpha}\mathcal{R}(u^\dagger)$ *such that*

$$w(u^\dagger) - w(u) \leq \beta_1 D_w^{\text{poly}}(u; u^\dagger) + \beta_2 \|K(u) - v^\dagger\| \tag{13}$$

holds for all u *with* $\mathcal{T}_{\bar{\alpha}}(u; v^\dagger) \leq \rho$.

If $q > 1$, *assume* $\alpha(\delta) \sim \delta^{q-1}$. *Then*

$$D_w^{\text{poly}}(u_\alpha^\delta; u^\dagger) = O(\delta) \quad \textit{and} \quad \|K(u_\alpha^\delta) - v^\delta\| = O(\delta).$$

If $q = 1$, *assume* $\alpha(\delta) \sim \delta^\epsilon$ *for* $\epsilon \in (0,1)$. *Then*

$$D_w^{\text{poly}}(u_\alpha^\delta; u^\dagger) = O(\delta^{1-\epsilon}) \quad \textit{and} \quad \|K(u_\alpha^\delta) - v^\delta\| = O(\delta).$$

Proof. This proof is analogous to the one of Proposition 3.41 in [19]. We include it here for the sake of completeness.

Since u_α^δ minimizes $\mathcal{T}_\alpha(\cdot; v^\delta)$ and $\|v^\dagger - v^\delta\| \leq \delta$ we have

$$\mathcal{T}_\alpha(u_\alpha^\delta; v^\delta) \leq \mathcal{T}_\alpha(u^\dagger; v^\delta) \leq \delta^q + \alpha\mathcal{R}(u^\dagger).$$

Adding $\alpha(D_w^{\text{poly}}(u_\alpha^\delta; u^\dagger) - \mathcal{R}(u_\alpha^\delta))$ to the left and right hand sides of this inequality gives

$$\begin{aligned} \|K(u_\alpha^\delta) - v^\delta\|^q + \alpha D_w^{\text{poly}}(u_\alpha^\delta; u^\dagger) &\leq \delta^q + \alpha(\mathcal{R}(u^\dagger) + D_w^{\text{poly}}(u_\alpha^\delta; u^\dagger) - \mathcal{R}(u_\alpha^\delta)) \\ &= \delta^q + \alpha(w(u^\dagger) - w(u_\alpha^\delta)). \end{aligned} \tag{14}$$

Now, let $\bar{\alpha} > 0$ and $\rho > \bar{\alpha}\mathcal{R}(u^\dagger)$. Remark 3.27 in [19] shows that $\mathcal{T}_{\bar{\alpha}}(u_\alpha^\delta; v^\dagger) \leq \rho$ for δ sufficiently small. (The possible lack of convexity of $\mathcal{R}$ does not make a difference here.) Therefore, we can use (13) with $u = u_\alpha^\delta$ to further estimate the right hand side of (14) yielding

$$\|K(u_\alpha^\delta) - v^\delta\|^q + \alpha D_w^{\text{poly}}(u_\alpha^\delta; u^\dagger) \leq \delta^q + \alpha(\beta_1 D_w^{\text{poly}}(u_\alpha^\delta; u^\dagger) + \beta_2\|K(u_\alpha^\delta) - v^\dagger\|)$$

Exploiting once again the fact that $\|v^\dagger - v^\delta\| \le \delta$ we arrive at

$$\begin{aligned}\|K(u_\alpha^\delta) - v^\delta\|^q + \alpha D_w^{\mathrm{poly}}(u_\alpha^\delta; u^\dagger) \le \delta^q + \alpha\beta_1 D_w^{\mathrm{poly}}(u_\alpha^\delta; u^\dagger) \\ + \alpha\beta_2(\|K(u_\alpha^\delta) - v^\delta\| + \delta),\end{aligned} \tag{15}$$

which holds for all sufficiently small $\delta > 0$.

Assume first that $q = 1$. In this case we can rewrite inequality (15) in the following way

$$(1 - \alpha\beta_2)\|K(u_\alpha^\delta) - v^\delta\| + \alpha(1 - \beta_1) D_w^{\mathrm{poly}}(u_\alpha^\delta; u^\dagger) \le (1 + \alpha\beta_2)\delta,$$

which directly implies that

$$\begin{aligned} D_w^{\mathrm{poly}}(u_\alpha^\delta; u^\dagger) &\le \frac{(1 + \alpha\beta_2)}{\alpha(1 - \beta_1)}\delta, \\ \|K(u_\alpha^\delta) - v^\delta\| &\le \frac{(1 + \alpha\beta_2)}{(1 - \alpha\beta_2)}\delta. \end{aligned}$$

Note that $\beta_1 < 1$ and that $\alpha(\delta)\beta_2 < 1$ for δ sufficiently small. If the parameter choice rule satisfies $\alpha(\delta) \sim \delta^\epsilon$, $0 < \epsilon < 1$, the assertion follows.

Now let $q > 1$ and rewrite (15) as

$$\|K(u_\alpha^\delta) - v^\delta\|^q - \alpha\beta_2\|K(u_\alpha^\delta) - v^\delta\| + \alpha(1 - \beta_1) D_w^{\mathrm{poly}}(u_\alpha^\delta; u^\dagger) \le \delta^q + \alpha\beta_2\delta.$$

Using Young's inequality $ab \le a^q/q + b^{q^*}/q^*$ for $a = \|K(u_\alpha^\delta) - v^\delta\|$ and $b = \alpha\beta_2$ we obtain

$$\Big(1 - \frac{1}{q}\Big)\|K(u_\alpha^\delta) - v^\delta\|^q + \alpha(1 - \beta_1) D_w^{\mathrm{poly}}(u_\alpha^\delta; u^\dagger) \le \delta^q + \alpha\beta_2\delta + (\alpha\beta_2)^{q^*}/q^*,$$

and consequently

$$\begin{aligned} D_w^{\mathrm{poly}}(u_\alpha^\delta; u^\dagger) &\le \frac{\delta^q + \alpha\beta_2\delta + (\alpha\beta_2)^{q^*}/q^*}{\alpha(1 - \beta_1)}, \\ \|K(u_\alpha^\delta) - v^\delta\|^q &\le \frac{q}{q - 1}\Big(\delta^q + \alpha\beta_2\delta + (\alpha\beta_2)^{q^*}/q^*\Big). \end{aligned}$$

For $\alpha(\delta) \sim \delta^{q-1}$ the assertion follows. □

Remark 4.5. Theorem 4.1 is a generalization of Theorem 2.2 in the following sense. Whenever Theorem 2.2 applies to a variational regularization method over $W^{1,p}(\Omega, \mathbb{R}^N)$, Theorem 4.1 applies as well with $w = (u^*, 0)$. In addition we have $D_w^{\mathrm{poly}}(u_\alpha^\delta; u^\dagger) = D_{u^*}(u_\alpha^\delta; u^\dagger)$.

Remark 4.6. On the other hand, the results of Theorem 4.1 can be seen as special cases of those in Section 3 of [14]. More specifically, if the assumptions of Theorem 4.1 hold, then $\mathcal{R}$ is locally W_{poly}-convex at $u^\dagger$ in the sense of [14] (cf. Remark 4.4) and minimization of $\mathcal{T}_\alpha$ satisfies a variational inequality at $u^\dagger$ (again in the sense of [14]).

Remark 4.7. Note that Theorem 4.1 does not require $\mathcal{R}$ to have a polyconvex integrand, just as Theorem 2.2 does not require $\mathcal{R}$ to be convex. However, a general non-polyconvex integrand cannot be expected to give rise to a W_{poly}-subdifferentiable functional $\mathcal{R}$.

Example 4.3. Let $\Omega \subset \mathbb{R}^n$ be a bounded open set with Lipschitz boundary and $n = 3$. Consider the volumetric regularization energy from [17], which can be written as

$$\mathcal{R}(u) = E_{\text{vol}}(u) = \int_\Omega F(\nabla u(x), \operatorname{adj}_2(\nabla u(x)), \det \nabla u(x))\, dx,$$

where

$$F(A,B,C) = \begin{cases} \eta_1 |A|^p + \eta_2 |B|^r + \eta_3 C^{-s}, & \text{if } C > 0, \\ +\infty, & \text{if } C \leq 0. \end{cases}$$

Here, the η_i are arbitrary positive constants whereas $p, r > n$ and $s > r(n-1)/(r-n)$. As in Example 3.1 it follows that minimization of

$$\mathcal{T}_\alpha(u; I_1^\delta) = \|I_2 \circ u - I_1^\delta\|^2_{L^2(\Omega)} + \alpha E_{\text{vol}}(u)$$

over $U = W^{1,p}(\Omega, \mathbb{R}^3)$ is a well-defined regularization method.

Next, assume $u^\dagger \in \operatorname{dom} E_{\text{vol}}$ is an $\mathcal{R}$-minimizing solution. According to Lemma 4.1 E_{vol} is W_{poly}-subdifferentiable at $u^\dagger$, if

$$|\operatorname{adj}_2(\nabla u^\dagger)|^{r-1} \in L^{(\frac{p}{2})^*} \quad \text{and} \quad |\det \nabla u^\dagger|^{-s-1} \in L^{(\frac{p}{3})^*},$$

and in this case Lemma 4.1 also provides an explicit formula for a W_{poly}-subgradient. If now, in addition, the source condition (13) is satisfied at $u^\dagger$, then we obtain a linear convergence rate in the W_{poly}-Bregman distance. However, finding specific examples where this is the case is non-trivial and remains an open question.

5 Conclusion

Convexity is an unnecessarily strong requirement for functionals $\mathcal{R}$ defined on $W^{1,p}(\Omega, \mathbb{R}^N)$, if the main concern is to ensure weak lower semicontinuity. In fact, polyconvexity of the integrand, or even quasiconvexity, is enough. However, if $\mathcal{R}$ is supposed to serve as a regularization functional, then the problem is how to measure convergence rates. The standard approach using classical Bregman distances $D_{u^*}(u_\alpha^\delta; u^\dagger)$ must be expected to fail in general due to the lack of convexity. In this article we have tried to answer two questions. (i) Are there instances of nonconvex variational regularization where standard convergence rates results *do* apply? (ii) What could a general strategy for obtaining convergence rates for regularization functionals with polyconvex integrands look like?

With Example 3.2 we have given a positive answer to the first question. It is based on the fact, explained in Remark 2.3, that source conditions are automatically satisfied, if $\mathcal{R}$ has a minimizer which is also an exact solution of the operator equation (6). Exploiting the fact that polyconvexity is compatible with minimality on $SO(n)$ we constructed an instance of the image registration problem with nonconvex regularization where a standard convergence rates result as given in Theorem 2.2 applies.

The second question was addressed by introducing W_{poly}-Bregman distances, which are based on a recent idea from [14] and which have a reasonably large domain of definition for a certain class of functionals with polyconvex integrands (cf. Lemma 4.1). By adapting the usual source conditions one can obtain linear convergence rates also for the W_{poly}-Bregman distances (see Theorem 4.1).

Acknowledgements

We thank José A. Iglesias for helpful comments and discussions, in particular regarding Section 3. Both authors acknowledge support by the Austrian Science Fund (FWF) within the national research network Geometry and Simulation, project S11704 (Variational Methods for Imaging on Manifolds). In addition, the work of OS is supported by the Austrian Science Fund (FWF), project P26687-N25 Interdisciplinary Coupled Physics Imaging.

References

[1] R. A. Adams and Fournier J. J. F. *Sobolev Spaces*. Pure and Applied Mathematics. Elsevier, Amsterdam, 2 edition, 2003.

[2] J. M. Ball. Convexity conditions and existence theorems in nonlinear elasticity. *Arch. Ration. Mech. Anal.*, 63:337–403, 1977.

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

[4] M. Burger, J. Modersitzki, and L. Ruthotto. A hyperelastic regularization energy for image registration. *SIAM J. Sci. Comput.*, 35(1):B132–B148, 2013.

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

[6] B. Dacorogna. *Direct methods in the calculus of variations*, volume 78 of *Applied Mathematical Sciences*. Springer, New York, second edition, 2008.

[7] M. Droske and M. Rumpf. A variational approach to non-rigid morphological registration. *SIAM J. Appl. Math.*, 64(2):668–687, 2004.

[8] I. Ekeland and R. Témam. *Convex analysis and variational problems*, volume 28 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, english edition, 1999. Translated from the French.

[9] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*. Number 375 in Mathematics and its Applications. Kluwer Academic Publishers Group, Dordrecht, 1996.

[10] B. Fischer and J. Modersitzki. Ill-posed medicine—an introduction to image registration. *Inverse Probl.*, 24(3):034008, 16, 2008.

[11] J. Flemming. Theory and examples of variational regularization with non-metric fitting functionals. *J. Inverse Ill-Posed Probl.*, 18(6):677–699, 2010.

[12] J. Flemming and B. Hofmann. A new approach to source conditions in regularization with general residual term. *Numer. Funct. Anal. Optim.*, 31(3):245–284, 2010.

[13] J. Flemming, B. Hofmann, and P. Mathé. Sharp converse results for the regularization error using distance functions. *Inverse Probl.*, 27:025006, 2011.

[14] M. Grasmair. Generalized Bregman distances and convergence rates for non-convex regularization methods. *Inverse Probl.*, 26(11):115014, October 2010.

[15] C. W. Groetsch. *Inverse problems in the mathematical sciences*. Vieweg Mathematics for Scientists and Engineers. Friedr. Vieweg & Sohn, Braunschweig, 1993.

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

[17] J. A. Iglesias, M. Rumpf, and O. Scherzer. Shape aware matching of implicit surfaces based on thin shell energies. Preprint on ArXiv arXiv:1509.06559, University of Vienna, Austria, 2015. Funded by the Austrian Science Fund (FWF) within the FSP S117 - "Geometry and Simulation".

[18] P. Pedregal. *Variational methods in nonlinear elasticity*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000.

[19] 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.

[20] T. Schuster, B. Kaltenbacher, B. Hofmann, and K. S. Kazimierski. *Regularization methods in Banach spaces*. Number 10 in Radon Series on Computational and Applied Mathematics. De Gruyter, Berlin, Boston, 2012.

[21] I. Singer. *Abstract convex analysis*. Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley & Sons Inc., New York, 1997. With a foreword by A. M. Rubinov, A Wiley-Interscience Publication.

[22] A. Sotiras, C. Davatzikos, and N. Paragios. Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190, July 2013.

[23] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems*. John Wiley & Sons, Washington, D.C., 1977.

[24] A. N. Tikhonov, A. Goncharsky, V. Stepanov, and A. Yagola. *Numerical Methods for the Solution of Ill-Posed Problems*. Kluwer, Dordrecht, 1995.

[25] A. N. Tikhonov, A. S. Leonov, and A. G. Yagola. *Nonlinear ill-posed problems. Vol. 1, 2*, volume 14 of *Applied Mathematics and Mathematical Computation*. Chapman & Hall, London, 1998. Translated from the Russian.

[26] C. A. Zarzer. On Tikhonov regularization with non-convex sparsity constraints. *Inverse Probl.*, 25:025006, 2009.