

# **A range condition for polyconvex variational regularization**

**C. Kirisits, O. Scherzer**

**RICAM-Report 2018-04**

# A range condition for polyconvex variational regularization

Clemens Kirisits\*, Otmar Scherzer†

February 14, 2018

## Abstract

In the context of *convex* variational regularization it is a known result that, under suitable differentiability assumptions, source conditions in the form of variational inequalities imply range conditions, while the converse implication only holds under an additional restriction on the operator. In this article we prove the analogous result for *polyconvex* regularization. More precisely, we show that the variational inequality derived in [10] implies that the derivative of the regularization functional must lie in the range of the dual-adjoint of the derivative of the operator. In addition, we show how to adapt the restriction on the operator in order to obtain the converse implication.

## 1 Introduction

Consider a nonlinear operator equation with inexact data

$$K(u) = v^\delta, \quad \|v^\delta - v^\dagger\| \leq \delta,$$

where  $K : U \rightarrow V$  acts between Banach spaces,  $v^\dagger, v^\delta \in V$  are exact and noisy data, respectively, and  $\delta > 0$  is the noise level. A common method for the stable inversion of  $K$  is variational regularization which consists in computing regularized solutions  $u_\alpha^\delta$  as minimizers of functionals of the form

$$u \mapsto \mathcal{T}_\alpha(u; v^\delta) = \|K(u) - v^\delta\|^q + \alpha \mathcal{R}(u). \quad (1)$$

Here  $\mathcal{R}$  is a typically convex regularization functional,  $\alpha > 0$  and  $q \geq 1$ . A natural requirement for such methods is that regularized solutions converge, in some sense, to an exact solution as the noise level tends to zero. Convergence

---

\*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

rates additionally provide bounds on the discrepancy between regularized and exact solutions in terms of the noise level. In a Banach space setting the most common measure of discrepancy is the Bregman distance associated to  $\mathcal{R}$  [4].

In order to guarantee convergence rates one has to impose a source condition of some sort. Traditionally, in a linear Hilbert space setting with quadratic Tikhonov regularization, this was done by assuming that the minimum norm solution lies in the range of an operator closely related to the adjoint of  $K$ . See [6, Ch. 5] for example. Generalizing this range condition to the nonlinear Banach space setting outlined in the previous paragraph yields

$$\mathcal{R}'(u^\dagger) \in \text{ran } K'(u^\dagger)^\#, \quad (2)$$

where  $u^\dagger$  is an  $\mathcal{R}$ -minimizing solution and  $K'(u^\dagger)^\#$  is the dual-adjoint of the Gâteaux derivative of  $K$  at  $u^\dagger$ .

More recently, it was shown in [8] that convergence rates can also be obtained by assuming that a variational inequality like

$$\langle u^*, u^\dagger - u \rangle \leq \beta_1 D_{u^*}(u; u^\dagger) + \beta_2 \|K(u) - v^\dagger\| \quad (3)$$

holds for all  $u$  in a certain neighbourhood of  $u^\dagger$ . Here  $u^*$  is a subgradient of  $\mathcal{R}$  at  $u^\dagger$  and  $D_{u^*}(u; u^\dagger)$  denotes the corresponding Bregman distance between  $u$  and  $u^\dagger$ . Note that (3) does not require  $K$  or  $\mathcal{R}$  to be differentiable. If they are, however, then the variational inequality (3) implies the range condition (2). The converse implication only holds under an additional assumption on the nonlinearity of the operator  $K$ . For a more detailed discussion on the relations between the various types of source conditions we refer to [11, pp. 70–73].

For certain inverse problems on  $W^{1,p}(\Omega, \mathbb{R}^N)$ , such as image or shape registration models inspired by nonlinear elasticity [3, 9], convex regularization is too restrictive, while the weaker notion of polyconvexity is more appropriate. Indeed, nonconvex regularization functionals  $\mathcal{R}$  with polyconvex integrands are well-suited for deriving stable and convergent regularization schemes. However, since such functionals are not subdifferentiable in general, the question is how to obtain convergence rates. In [10] we addressed this issue by following Grasmair’s approach of generalized Bregman distances [7]. First, we introduced the weaker concept of  $W_{\text{poly}}$ -subdifferentiability, specifically designed for functionals with polyconvex integrands, and gave conditions for existence of  $W_{\text{poly}}$ -subgradients. By means of the corresponding  $W_{\text{poly}}$ -Bregman distance we were then able to translate the convergence rates result of [8] to the polyconvex setting. The source condition derived in [10] reads

$$w(u^\dagger) - w(u) \leq \beta_1 D_w^{\text{poly}}(u; u^\dagger) + \beta_2 \|K(u) - v^\dagger\|, \quad (4)$$

where  $w$  is a  $W_{\text{poly}}$ -subgradient of  $\mathcal{R}$  at  $u^\dagger$  and  $D_w^{\text{poly}}(u; u^\dagger)$  is the corresponding generalized Bregman distance.

The main results of the present paper are Theorems 1 and 2 in Section 3. Theorem 1 states that the variational inequality (4) implies the range condition

(2), given that  $K$  and  $\mathcal{R}$  are differentiable and  $\mathcal{R}$  satisfies the conditions guaranteeing existence of a  $W_{\text{poly}}$ -subgradient. A major part of the proof consists in showing that  $\mathcal{R}'(u^\dagger) = w'(u^\dagger)$  in this case. Conversely, Theorem 2 states that

$$w'(u^\dagger) \in \text{ran } K'(u^\dagger)^\#$$

implies (4), if the nonlinearities of  $K$  and  $w$  satisfy a certain inequality around  $u^\dagger$ .

## 2 Polyconvex functions and generalized Bregman distances

This section is a brief summary of the most important prerequisites from [10]. For  $N, n \in \mathbb{N}$  we will frequently identify matrices in  $\mathbb{R}^{N \times n}$  with vectors in  $\mathbb{R}^{Nn}$ .

**Polyconvex functions.** A function  $f : \mathbb{R}^{N \times n} \rightarrow \mathbb{R} \cup \{+\infty\}$  is *polyconvex*, if, for every  $A \in \mathbb{R}^{N \times n}$ ,  $f(A)$  can be written as a convex function of the minors of  $A$ . More precisely, let  $1 \leq s \leq \min(N, n) =: N \wedge n$  and define  $\sigma(s) := \binom{n}{s} \binom{N}{s}$  as well as  $\tau := \sum_{s=1}^{N \wedge n} \sigma(s)$ . Denote by  $\text{adj}_s A \in \mathbb{R}^{\sigma(s)}$  the matrix of all  $s \times s$  minors of  $A$  and set

$$T(A) := (A, \text{adj}_2 A, \dots, \text{adj}_{N \wedge n} A) \in \mathbb{R}^\tau.$$

Now, a function  $f : \mathbb{R}^{N \times n} \rightarrow \mathbb{R} \cup \{+\infty\}$  is polyconvex, if there is a convex function  $F : \mathbb{R}^\tau \rightarrow \mathbb{R} \cup \{+\infty\}$  such that  $f = F \circ T$ . Every convex function is polyconvex. The converse statement only holds, if  $N \wedge n = 1$ . The importance of polyconvex functions in the calculus of variations is due to the fact that they render functionals of the form

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

weakly lower semicontinuous in  $W^{1,p}(\Omega, \mathbb{R}^N)$ , where  $\Omega \subset \mathbb{R}^n$ . For more details on polyconvex functions see [1, 5].

**The set  $W_{\text{poly}}$ .** For the remainder of this article, unless stated otherwise, we let  $\Omega \subset \mathbb{R}^n$  be an open set,  $p \geq N \wedge n$ , and set  $U = W^{1,p}(\Omega, \mathbb{R}^N)$ .

The following variant of the map  $T$  will prove useful. Set  $\tau_2 := \sum_{s=2}^{N \wedge n} \sigma(s)$  and define

$$T_2(A) := (\text{adj}_2 A, \dots, \text{adj}_{N \wedge n} A) \in \mathbb{R}^{\tau_2}.$$

If  $u \in U$ , then  $\text{adj}_s \nabla u$  consists of sums of products of  $s$   $L^p(\Omega)$  functions, and therefore, by Hölder's inequality,  $\text{adj}_s \nabla u \in L^{p/s}(\Omega, \mathbb{R}^{\sigma(s)})$ . This motivates the following two definitions

$$S := \prod_{s=1}^{N \wedge n} L^{\frac{p}{s}}(\Omega, \mathbb{R}^{\sigma(s)}), \quad S_2 := \prod_{s=2}^{N \wedge n} L^{\frac{p}{s}}(\Omega, \mathbb{R}^{\sigma(s)}).$$

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

$$w(u) = \langle u^*, u \rangle_{U^*, U} + \langle v^*, T_2(\nabla u) \rangle_{S_2^*, S_2} \quad (5)$$

for all  $u \in U$ . Note that, if  $v^* = 0$ , then  $w$  can be identified with  $u^* \in U^*$ . Thus, the dual  $U^*$  can be regarded a subset of  $W_{\text{poly}}$  in a natural way.

**Generalized subgradients.** Let  $\mathcal{R} : U \rightarrow \mathbb{R} \cup \{+\infty\}$ . We denote the effective domain of  $\mathcal{R}$  by  $\text{dom } \mathcal{R} = \{u \in U : \mathcal{R}(u) < +\infty\}$ . Following [7, 10, 12] we define the  $W_{\text{poly}}$ -subdifferential of  $\mathcal{R}$  at  $u \in \text{dom } \mathcal{R}$  as

$$\partial_{\text{poly}} \mathcal{R}(u) = \{w \in W_{\text{poly}} : \mathcal{R}(v) \geq \mathcal{R}(u) + w(v) - w(u) \text{ for all } v \in U\},$$

If  $\mathcal{R}(u) = +\infty$ , we set  $\partial_{\text{poly}} \mathcal{R}(u) = \emptyset$ . The identification of  $U^*$  with a subset of  $W_{\text{poly}}$  mentioned in the previous paragraph implies that  $\partial \mathcal{R}(u) \subset \partial_{\text{poly}} \mathcal{R}(u)$ , that is, the classical subdifferential can be regarded a subset of the  $W_{\text{poly}}$ -subdifferential. Elements of  $\partial_{\text{poly}} \mathcal{R}(u)$  are called  $W_{\text{poly}}$ -subgradients of  $\mathcal{R}$  at  $u$ . Concerning existence of  $W_{\text{poly}}$ -subgradients we have shown the following result in [10].

**Lemma 1.** *Let*

$$F : \Omega \times \mathbb{R}^N \times \mathbb{R}^\tau \rightarrow \mathbb{R}_{\geq 0} \cup \{+\infty\}$$

*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 and denote its derivative by  $F'_{u, \xi}$ . 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  $x \mapsto F'_{u, \xi}(x, \bar{v}(x), T(\nabla \bar{v}(x)))$  lies in  $L^{p^*}(\Omega, \mathbb{R}^N) \times S^*$ , where  $p^*$  denotes the Hölder conjugate of  $p$ , then this function is a  $W_{\text{poly}}$ -subgradient of  $\mathcal{R}$  at  $\bar{v}$ .*

**Remark 1.** If  $F'_{u, \xi}(\cdot, \bar{v}(\cdot), T(\nabla \bar{v}(\cdot)))$  is a  $W_{\text{poly}}$ -subgradient  $w \in \partial_{\text{poly}} \mathcal{R}(\bar{v}) \subset W_{\text{poly}}$ , as postulated by Lemma 1, then it must be possible to write its action on  $u \in U$  in terms of a pair  $(u^*, v^*) \in U^* \times S_2^*$  as in (5). In order to do so recall that  $T(A) = (A, T_2(A))$ . We can split the variable  $\xi \in \mathbb{R}^\tau$  accordingly into  $(\xi_1, \xi_2) \in \mathbb{R}^{Nn} \times \mathbb{R}^{\tau_2}$ . Similarly, we can write  $F'_{u, \xi} = (F'_u, F'_\xi) = (F'_u, F'_{\xi_1}, F'_{\xi_2})$ . Now we have

$$\begin{aligned} w(u) &= \int_{\Omega} F'_{u, \xi}(x, \bar{v}(x), T(\nabla \bar{v}(x))) \cdot (u, T(\nabla u)) dx \\ &= \int_{\Omega} F'_u(x, \bar{v}(x), T(\nabla \bar{v}(x))) \cdot u(x) dx \\ &\quad + \int_{\Omega} F'_{\xi_1}(x, \bar{v}(x), T(\nabla \bar{v}(x))) \cdot \nabla u(x) dx \\ &\quad + \int_{\Omega} F'_{\xi_2}(x, \bar{v}(x), T(\nabla \bar{v}(x))) \cdot T_2(\nabla u(x)) dx. \end{aligned}$$

The integral in the bottom line corresponds to the dual pairing  $\langle v^*, T_2(\nabla u) \rangle_{S_2^*, S_2}$  in (5), while the previous two terms correspond to  $\langle u^*, u \rangle_{U^*, U}$ . Therefore,  $u^*$  is given by  $(F'_u, F'_{\xi_1})$  and  $v^*$  by  $F'_{\xi_2}$ . Also note that all integrals are well-defined and finite because of the integrability conditions on the derivative of  $F$  in Lemma 1.

**Generalized Bregman distances.** Whenever  $\mathcal{R}$  has a  $W_{\text{poly}}$ -subgradient  $w \in \partial_{\text{poly}}\mathcal{R}(u)$  we can define the associated  $W_{\text{poly}}$ -Bregman distance between  $v \in U$  and  $u$  as

$$D_w^{\text{poly}}(v; u) = \mathcal{R}(v) - \mathcal{R}(u) - w(v) + w(u).$$

Note that, just like the classical Bregman distance, the  $W_{\text{poly}}$ -Bregman distance is nonnegative, satisfies  $D_w^{\text{poly}}(u; u) = 0$  whenever defined, and is *not* symmetric with respect to  $u$  and  $v$ . In addition, if  $w = (u^*, 0) \in \mathcal{R}_{\text{poly}}(u)$ , then  $u^* \in \partial\mathcal{R}(u)$  and the classical and  $W_{\text{poly}}$ -Bregman distances coincide, that is,

$$D_w^{\text{poly}}(v; u) = D_{u^*}(v; u).$$

See [7, 11] for more details on (generalized) Bregman distances.

In order to be able to quote the source condition from [10] we need one more definition: We call  $u^\dagger \in U$  an  $\mathcal{R}$ -minimizing solution, if it solves the exact operator equation and minimizes  $\mathcal{R}$  among all other exact solutions, that is,

$$u^\dagger \in \arg \min \{ \mathcal{R}(u) : u \in U, K(u) = v^\dagger \}.$$

**Assumption 1.** Assume that  $\mathcal{R}$  has a  $W_{\text{poly}}$ -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

$$w(u^\dagger) - w(u) \leq \beta_1 D_w^{\text{poly}}(u; u^\dagger) + \beta_2 \|K(u) - v^\dagger\| \quad (6)$$

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

### 3 A range condition

At the end of this section we prove our main results, Theorems 1 and 2. Before that we have to state a few preliminary results. First, we recall the definition of the dual-adjoint operator together with a characterization of its range (Lemma 2). Next, we compute the Gâteaux derivative of

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

in Lemma 3, and of  $w \in W_{\text{poly}}$  in Lemma 4, respectively.

For every bounded linear operator  $A : U \rightarrow V$  acting between locally convex spaces there exists a unique operator  $A^\# : V^* \rightarrow U^*$ , also linear and bounded, satisfying

$$\langle A^\# v^*, u \rangle_{U^*, U} = \langle v^*, Au \rangle_{V^*, V}$$

for all  $u \in U$  and  $v^* \in V^*$ . See, for instance, Section VII.1 of [13]. The operator  $A^\#$  is called the *dual-adjoint* of  $A$ .

**Lemma 2.** *Let  $U, V$  be normed linear spaces,  $A : U \rightarrow V$  a bounded linear operator and  $u^* \in U^*$ . Then  $u^* \in \text{ran } A^\#$ , if and only if there is a  $C > 0$  such that*

$$|\langle u^*, u \rangle_{U^*, U}| \leq C \|Au\|$$

for all  $u \in U$ .

*Proof.* See Lemma 8.21 in [11]. □

Let  $K : \mathcal{D}(K) \subset U \rightarrow V$  be a map acting between normed spaces and let  $u \in \mathcal{D}(K)$ ,  $h \in U$ . If the limit

$$K'(u; h) = \lim_{t \rightarrow 0^+} \frac{1}{t} (K(u + th) - K(u))$$

exists in  $V$ , then  $K'(u; h)$  is called the *directional derivative* of  $K$  at  $u$  in direction  $h$ . If  $K'(u; h)$  exists for all  $h \in U$  and there is a bounded linear operator  $K'(u) : U \rightarrow V$  satisfying

$$K'(u)h = K'(u; h)$$

for all  $h \in U$ , then  $K$  is *Gâteaux differentiable* at  $u$  and  $K'(u)$  is called the *Gâteaux derivative* of  $K$  at  $u$ .

**Lemma 3.** *Let*

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

*be a nonnegative Carathéodory function. Assume that, for almost every  $x \in \Omega$ , the map  $(u, A) \mapsto f(x, u, A)$  is differentiable throughout its effective domain and that*

$$|f'_{u,A}(x, u, A)| \leq a(x) + b|u|^{p-1} + c|A|^{p-1} \quad (7)$$

*holds there for  $p \geq 1$  and some  $a \in L^{p^*}(\Omega)$  and  $b, c \geq 0$ . Then the functional*

$$\mathcal{R} : U = W^{1,p}(\Omega, \mathbb{R}^N) \rightarrow \mathbb{R}_{\geq 0} \cup \{+\infty\},$$

*defined by*

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

*is Gâteaux differentiable in the interior of its effective domain. Its Gâteaux derivative at  $u \in \text{int dom } \mathcal{R}$  is given by*

$$\begin{aligned} \langle \mathcal{R}'(u), \hat{u} \rangle_{U^*, U} &= \int_{\Omega} f'_u(x, u(x), \nabla u(x)) \cdot \hat{u}(x) \, dx \\ &+ \int_{\Omega} f'_A(x, u(x), \nabla u(x)) \cdot \nabla \hat{u}(x) \, dx, \quad \hat{u} \in U. \end{aligned} \quad (8)$$

*Proof.* Fix  $u \in \text{int dom } \mathcal{R}$  and  $\hat{u} \in U$ . Assuming we can differentiate under the integral sign we have

$$\begin{aligned} \mathcal{R}'(u; \hat{u}) &= \lim_{t \rightarrow 0^+} \frac{1}{t} (\mathcal{R}(u + t\hat{u}) - \mathcal{R}(u)) \\ &= \int_{\Omega} \lim_{t \rightarrow 0^+} \frac{1}{t} (f(x, u + t\hat{u}, \nabla u + t\nabla\hat{u}) - f(x, u, \nabla u)) dx \\ &= \int_{\Omega} \partial_t f(x, u + t\hat{u}, \nabla u + t\nabla\hat{u}) \Big|_{t=0} dx \\ &= \int_{\Omega} (f'_u(x, u, \nabla u) \cdot \hat{u} + f'_A(x, u, \nabla u) \cdot \nabla\hat{u}) dx, \end{aligned}$$

which is just (8).

It remains to show that differentiation and integration are interchangeable. For  $\epsilon > 0$  sufficiently small (see below) we define  $g : (-\epsilon, \epsilon) \times \Omega \rightarrow \mathbb{R}_{\geq 0} \cup \{+\infty\}$ ,

$$g(t, x) = f(x, u(x) + t\hat{u}(x), \nabla u(x) + t\nabla\hat{u}(x)).$$

The identity  $\partial_t \int_{\Omega} g(t, x) dx = \int_{\Omega} \partial_t g(t, x) dx$  holds true, if the following three conditions are satisfied.

1. Integrability: The function  $x \mapsto g(t, x)$  is integrable for all  $t \in (-\epsilon, \epsilon)$ .
2. Differentiability: The partial derivative  $\partial_t g(t, x)$  exists for almost every  $x \in \Omega$  and all  $t \in (-\epsilon, \epsilon)$ .
3. Uniform upper bound: There is a function  $h \in L^1(\Omega)$  such that  $|\partial_t g(t, x)| \leq h(x)$  for almost every  $x \in \Omega$  and all  $t \in (-\epsilon, \epsilon)$ .

Item 1 is satisfied, since  $u$  lies in the interior of  $\text{dom } \mathcal{R}$  and therefore

$$\int_{\Omega} |g(t, x)| dx = \mathcal{R}(u + t\hat{u}) < \infty, \quad -\epsilon < t < \epsilon,$$

for  $\epsilon$  sufficiently small. In particular,  $g(t, x) < \infty$  for almost every  $x$  and every  $t \in (-\epsilon, \epsilon)$ . Thus, item 2 holds as well. Concerning item 3, we use assumption (7) to obtain for almost every  $x \in \Omega$

$$\begin{aligned} |\partial_t g(t, x)| &= |f'_u(x, u + t\hat{u}, \nabla u + t\nabla\hat{u}) \cdot \hat{u} + f'_A(x, u + t\hat{u}, \nabla u + t\nabla\hat{u}) \cdot \nabla\hat{u}| \\ &\leq |f'_u(x, u + t\hat{u}, \nabla u + t\nabla\hat{u})| |\hat{u}| + |f'_A(x, u + t\hat{u}, \nabla u + t\nabla\hat{u})| |\nabla\hat{u}| \\ &\leq (|\hat{u}| + |\nabla\hat{u}|)(a + b|u + t\hat{u}|^{p-1} + c|\nabla u + t\nabla\hat{u}|^{p-1}). \end{aligned}$$

We estimate further

$$|u + t\hat{u}|^{p-1} \leq (|u| + |t||\hat{u}|)^{p-1} \leq \max\{1, 2^{p-2}\} (|u|^{p-1} + \epsilon^{p-1} |\hat{u}|^{p-1})$$

and similarly

$$|\nabla u + t\nabla\hat{u}|^{p-1} \leq \max\{1, 2^{p-2}\} (|\nabla u|^{p-1} + \epsilon^{p-1} |\nabla\hat{u}|^{p-1}).$$

Thus we have found an upper bound for  $|\partial_t g(t, x)|$ , which is independent of  $t$ . This bound is essentially a sum of products of the form  $y(x)z(x)^{p-1}$ , where  $y, z \in L^p(\Omega)$ . Since, in this case,  $z^{p-1}$  lies in  $L^{p^*}(\Omega)$ , Hölder's inequality shows that  $yz^{p-1} \in L^1(\Omega)$ .  $\square$

**Lemma 4.** *The functions  $w \in W_{\text{poly}}$  are Gâteaux differentiable on all of  $U$ . Identifying  $w$  with  $(u^*, v^*) \in U^* \times S_2^*$  its Gâteaux derivative at  $u \in U$  is given by*

$$\langle w'(u), \hat{u} \rangle_{U^*, U} = \langle u^*, \hat{u} \rangle_{U^*, U} + \int_{\Omega} v^*(x) \cdot T_2'(\nabla u(x)) \nabla \hat{u}(x) dx, \quad \hat{u} \in U,$$

where  $T_2'(\nabla u(x))$  denotes the derivative of the map  $T_2 : \mathbb{R}^{Nn} \rightarrow \mathbb{R}^{\tau_2}$  at  $\nabla u(x)$ .

*Proof.* Identify  $w \in W_{\text{poly}}$  with  $(u^*, v^*) \in U^* \times S_2^*$  and let  $u, \hat{u} \in U$ . First, we separate the linear and nonlinear parts of  $w$ .

$$\begin{aligned} w'(u; \hat{u}) &= \lim_{t \rightarrow 0^+} \frac{1}{t} (w(u + t\hat{u}) - w(u)) \\ &= \lim_{t \rightarrow 0^+} \frac{1}{t} (\langle u^*, u + t\hat{u} \rangle_{U^*, U} + \langle v^*, T_2(\nabla u + t\nabla \hat{u}) \rangle_{S_2^*, S_2} \\ &\quad - \langle u^*, u \rangle_{U^*, U} - \langle v^*, T_2(\nabla u) \rangle_{S_2^*, S_2}) \\ &= \langle u^*, \hat{u} \rangle_{U^*, U} + \lim_{t \rightarrow 0^+} \frac{1}{t} \langle v^*, T_2(\nabla u + t\nabla \hat{u}) - T_2(\nabla u) \rangle_{S_2^*, S_2} \end{aligned}$$

Assuming we can differentiate under the integral sign, the remaining limit equals

$$\begin{aligned} &\lim_{t \rightarrow 0^+} \frac{1}{t} \langle v^*, T_2(\nabla u + t\nabla \hat{u}) - T_2(\nabla u) \rangle_{S_2^*, S_2} \\ &= \int_{\Omega} \lim_{t \rightarrow 0^+} \frac{1}{t} \left[ v^* \cdot (T_2(\nabla u + t\nabla \hat{u}) - T_2(\nabla u)) \right] dx \\ &= \int_{\Omega} \partial_t \left[ v^* \cdot T_2(\nabla u + t\nabla \hat{u}) \right]_{t=0} dx \\ &= \int_{\Omega} v^* \cdot T_2'(\nabla u) \nabla \hat{u} dx. \end{aligned}$$

As in the proof of Lemma 3 we have to check the conditions for interchanging integration and differentiation. Define the function

$$g(t, x) = v^*(x) \cdot T_2(\nabla u(x) + t\nabla \hat{u}(x))$$

on  $(-\epsilon, \epsilon) \times \Omega$ . It is integrable for all  $t$ , since  $T_2$  maps  $L^p(\Omega, \mathbb{R}^{N \times n})$  into  $S_2$  and  $v^*$  lies in  $S_2^*$ . It is also differentiable with respect to  $t$ , since the entries of  $T_2(\nabla u(x) + t\nabla \hat{u}(x))$  are polynomials in  $t$ . Finally,  $\partial_t g$  can be bounded in the

following way

$$\begin{aligned}
|\partial_t g| &= \left| \partial_t \sum_{s=2}^n v_s^* \cdot \text{adj}_s(\nabla u + t\nabla \hat{u}) \right| \\
&= \left| \sum_{s=2}^n v_s^* \cdot \text{adj}'_s(\nabla u + t\nabla \hat{u}) \nabla \hat{u} \right| \\
&\leq |\nabla \hat{u}| \sum_{s=2}^n |v_s^*| |\text{adj}'_s(\nabla u + t\nabla \hat{u})| \tag{9}
\end{aligned}$$

where  $v_s^*$  denotes the  $L^{(\frac{p}{s})^*}(\Omega, \mathbb{R}^{\sigma(s)})$ -component of  $v^*$ . The derivative  $\text{adj}'_s(\nabla u + t\nabla \hat{u})$  consists of sums of products of  $s-1$  terms of the form  $\partial_{x_i} u_j + t\partial_{x_i} \hat{u}_j$ . After expanding, every such product can be bounded by

$$\sum_{k=0}^{s-1} |t|^k \sum_m |g_{km}| \leq \sum_{k=0}^{s-1} \epsilon^k \sum_m |g_{km}|, \tag{10}$$

where each  $g_{km}$  is a product of  $s-1$   $L^p$  functions and therefore lies in  $L^{\frac{p}{s-1}}$ . Combining (9) with (10) gives an upper bound for  $\partial_t g$  which is independent of  $t$ . Using Hölder's inequality it is now straightforward to verify that this bound is indeed an  $L^1$  function.  $\square$

**Theorem 1.** *Let  $\mathcal{R}$  satisfy the requirements of Lemma 1 at an  $\mathcal{R}$ -minimizing solution  $u^\dagger \in \text{int dom } \mathcal{R}$  and let  $w$  be the  $W_{\text{poly}}$ -subgradient thus provided. Suppose Assumption 1 holds for this  $u^\dagger$  and  $w$ . Moreover, assume that the integrand  $f$  of  $\mathcal{R}$  satisfies inequality (7) and that  $K$  is Gâteaux differentiable at  $u^\dagger$ . Then  $\mathcal{R}$  is Gâteaux differentiable at  $u^\dagger$  and*

$$\mathcal{R}'(u^\dagger) = w'(u^\dagger) \in \text{ran } K'(u^\dagger)^\#.$$

*Proof.* The proof consists of two steps. First, we show that the source condition implies that

$$0 \leq \beta_1 \langle \mathcal{R}'(u^\dagger), \hat{u} \rangle_{U^*, U} + (1 - \beta_1) \langle w'(u^\dagger), \hat{u} \rangle_{U^*, U} + \beta_2 \|K'(u^\dagger)\hat{u}\| \tag{11}$$

holds for all  $\hat{u} \in U$ . Second, the derivatives of  $\mathcal{R}$  and  $w$  at  $u^\dagger$  agree, which leads to

$$\langle \mathcal{R}'(u^\dagger), \hat{u} \rangle_{U^*, U} \leq \beta_2 \|K'(u^\dagger)\hat{u}\|$$

for all  $\hat{u} \in U$ . Finally, Lemma 2 implies  $\mathcal{R}'(u^\dagger) \in \text{ran } K'(u^\dagger)^\#$ .

Step 1: Inequality (6) can be equivalently written as

$$0 \leq \beta_1 (\mathcal{R}(u) - \mathcal{R}(u^\dagger)) + (1 - \beta_1) (w(u) - w(u^\dagger)) + \beta_2 \|K(u) - K(u^\dagger)\|.$$

Since  $\mathcal{R}$  satisfies the requirements of Lemma 1 as well as inequality (7), Lemma 3 applies. Now, because of differentiability of both  $\mathcal{R}$  and  $K$  at  $u^\dagger$  and because

$\mathcal{T}_{\bar{\alpha}}(u^\dagger; v^\dagger) < \rho$  by Assumption 1, there is a  $t_0 > 0$  for every  $\hat{u} \in U$  such that  $\mathcal{T}_{\bar{\alpha}}(u^\dagger + t\hat{u}; v^\dagger) < \rho$  for  $0 \leq t < t_0$ . Therefore,

$$0 \leq \beta_1(\mathcal{R}(u^\dagger + t\hat{u}) - \mathcal{R}(u^\dagger)) + (1 - \beta_1)(w(u^\dagger + t\hat{u}) - w(u^\dagger)) + \beta_2 \|K(u^\dagger + t\hat{u}) - K(u^\dagger)\|.$$

Dividing by  $t$  and letting  $t \rightarrow 0$  yields (11).

Step 2: We now show that  $\mathcal{R}'(u^\dagger) = w'(u^\dagger)$ . By Lemma 3 the derivative of  $\mathcal{R}$  is given by

$$\langle \mathcal{R}'(u^\dagger), \hat{u} \rangle_{U^*, U} = \int_{\Omega} f'_u(x, u^\dagger, \nabla u^\dagger) \cdot \hat{u} \, dx + \int_{\Omega} f'_A(x, u^\dagger, \nabla u^\dagger) \cdot \nabla \hat{u} \, dx.$$

Since  $f(x, u, A) = F(x, u, T(A))$ , the chain rule yields

$$\begin{aligned} \langle \mathcal{R}'(u^\dagger), \hat{u} \rangle_{U^*, U} &= \int_{\Omega} F'_u(x, u^\dagger, T(\nabla u^\dagger)) \cdot \hat{u} \, dx \\ &\quad + \int_{\Omega} F'_\xi(x, u^\dagger, T(\nabla u^\dagger)) \cdot T'(\nabla u^\dagger) \nabla \hat{u} \, dx. \end{aligned}$$

Now we split  $F'_\xi$  into  $(F'_{\xi_1}, F'_{\xi_2})$  as in Remark 1 and, accordingly,  $T'(\nabla u^\dagger)$  into  $(\text{Id}, T'_2(\nabla u^\dagger))$  where  $\text{Id}$  is the identity mapping on  $\mathbb{R}^{Nn}$ . This leads to

$$\begin{aligned} \langle \mathcal{R}'(u^\dagger), \hat{u} \rangle_{U^*, U} &= \int_{\Omega} F'_u(x, u^\dagger, T(\nabla u^\dagger)) \cdot \hat{u} \, dx + \int_{\Omega} F'_{\xi_1}(x, u^\dagger, T(\nabla u^\dagger)) \cdot \nabla \hat{u} \, dx \\ &\quad + \int_{\Omega} F'_{\xi_2}(x, u^\dagger, T(\nabla u^\dagger)) \cdot T'_2(\nabla u^\dagger) \nabla \hat{u} \, dx. \end{aligned}$$

On the other hand, recall Remark 1 to see that the  $W_{\text{poly}}$ -subgradient  $w \in \partial_{\text{poly}} \mathcal{R}(u^\dagger)$  provided by Lemma 1 is given by

$$\begin{aligned} w(u) &= \underbrace{\int_{\Omega} F'_u(x, u^\dagger, T(\nabla u^\dagger)) \cdot u \, dx + \int_{\Omega} F'_{\xi_1}(x, u^\dagger, T(\nabla u^\dagger)) \cdot \nabla u \, dx}_{=\langle u^*, u \rangle_{U^*, U}} \\ &\quad + \underbrace{\int_{\Omega} F'_{\xi_2}(x, u^\dagger, T(\nabla u^\dagger)) \cdot T_2(\nabla u) \, dx}_{=\langle v^*, T_2(\nabla u) \rangle_{S_2^*, S_2}}. \end{aligned}$$

Computing the derivative of  $w$  according to Lemma 4 shows that  $\mathcal{R}'(u^\dagger) = w'(u^\dagger)$ .  $\square$

**Remark 2.** Theorem 1 is an extension of its counterpart from convex regularization theory, Proposition 3.38 in [11], in the following sense. If the latter applies to a variational regularization method on  $U$  with  $\mathcal{R}$  being as in Lemma 3 but convex, then Thm. 1 applies as well with  $w \in \partial_{\text{poly}} \mathcal{R}(u^\dagger)$  and  $D_w^{\text{poly}}(u; u^\dagger)$  reducing to their classical analogues and the respective variational inequalities and range conditions being identical. See also [10, Remark 4.5].

**Theorem 2.** *Assume  $K$  is Gâteaux differentiable at an  $\mathcal{R}$ -minimizing solution  $u^\dagger$  and that  $\mathcal{R}$  has a  $W_{\text{poly}}$ -subgradient  $w$  there. In addition, suppose there is a  $\omega^* \in V^*$  as well as constants  $\beta_1 \in [0, 1)$ ,  $\bar{\alpha} > 0$ ,  $\rho > \bar{\alpha}\mathcal{R}(u^\dagger)$  such that*

$$w'(u^\dagger) = K'(u^\dagger)\# \omega^*, \quad \text{and} \quad (12)$$

$$\begin{aligned} \|\omega^*\| \|K(u) - v^\dagger - K'(u^\dagger)(u - u^\dagger)\| + w(u^\dagger) - w(u) \\ - \langle w'(u^\dagger), u^\dagger - u \rangle_{U^*, U} \leq \beta_1 D_w^{\text{poly}}(u; u^\dagger) \end{aligned} \quad (13)$$

for all  $u$  satisfying  $\mathcal{T}_{\bar{\alpha}}(u; v^\dagger) \leq \rho$ . Then Assumption 1 holds.

*Proof.* The proof is along the lines of [11, Prop. 3.35]. We include it here in order to clarify the main differences.

By virtue of (12) we have for every  $u \in U$

$$\begin{aligned} \langle w'(u^\dagger), u^\dagger - u \rangle_{U^*, U} &= \langle K'(u^\dagger)\# \omega^*, u^\dagger - u \rangle_{U^*, U} \\ &= \langle \omega^*, K'(u^\dagger)(u^\dagger - u) \rangle_{U^*, U} \\ &= \|\omega^*\| \|K'(u^\dagger)(u^\dagger - u)\| \\ &\leq \|\omega^*\| \|K(u) - v^\dagger\| + \|\omega^*\| \|K(u) - v^\dagger - K'(u^\dagger)(u - u^\dagger)\|. \end{aligned}$$

Adding  $w(u^\dagger) - w(u) - \langle w'(u^\dagger), u^\dagger - u \rangle_{U^*, U}$  on both sides and using (13) we arrive at

$$w(u^\dagger) - w(u) \leq \|\omega^*\| \|K(u) - v^\dagger\| + \beta_1 D_w^{\text{poly}}(u; u^\dagger),$$

which is just (6) with  $\beta_2 = \|\omega^*\|$ .  $\square$

**Remark 3.** Note that the expression

$$w(u^\dagger) - w(u) - \langle w'(u^\dagger), u^\dagger - u \rangle_{U^*, U} \quad (14)$$

in (13) is just the difference between  $w$  and its continuous affine approximation around  $u^\dagger$ . Therefore, condition (13) is essentially a restriction on the nonlinearity of  $K$  plus the nonlinearity of  $w$ , both computed in a neighbourhood of  $u^\dagger$ .

Theorem 2 extends [11, Prop. 3.35] in the same way Theorem 1 extends [11, Prop. 3.38]. If  $w = (u^*, 0)$ , then  $w'(u^\dagger) = u^*$  and the nonlinearity (14) vanishes.

## 4 Conclusion

In recent years, several authors have shown that nonconvex regularization of inverse problems is not only a viable possibility, but can even be preferable to convex regularization in certain situations, see for instance [2, 3, 7, 9, 10, 14]. However, convergence rates results for nonconvex regularization are exceedingly rare, let alone results relating different types of source conditions.

In this paper we have shown that two such results can be translated to the polyconvex setting of [10]. The first one states that, under suitable differentiability assumptions, source conditions in the form of variational inequalities

imply range conditions. One of the reasons why this statement remains true is the fact that the derivative of  $\mathcal{R}$  is equal to the derivative of its  $W_{\text{poly}}$ -subgradient. This fact can be interpreted as a generalization of the well-known identity  $\partial\mathcal{R}(u) = \{\mathcal{R}'(u)\}$  for convex and differentiable functions  $\mathcal{R}$ . Second, we have demonstrated that a converse statement can be obtained as well, given that the sum of the nonlinearities of  $K$  and of the  $W_{\text{poly}}$ -subgradient can be bounded by the  $W_{\text{poly}}$ -Bregman distance around  $u^\dagger$ .

## Acknowledgements

Both authors acknowledge support by the Austrian Science Fund (FWF): S117. In addition, the work of OS is supported by the FWF Sonderforschungsbereich (SFB) F 68, as well as by project I 3661, jointly funded by FWF and Deutsche Forschungsgemeinschaft (DFG).

## References

- [1] J. M. Ball. Convexity conditions and existence theorems in nonlinear elasticity. *Arch. Ration. Mech. Anal.*, 63:337–403, 1977.
- [2] K. Bredies and D. Lorenz. Regularization with non-convex separable constraints. *Inverse Probl.*, 25(8):085011 (14pp), 2009.
- [3] M. Burger, J. Modersitzki, and L. Ruthotto. A hyperelastic regularization energy for image registration. *SIAM J. Sci. Comput.*, 35(1):B132–B148, 2013.
- [4] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Probl.*, 20(5):1411–1421, 2004.
- [5] B. Dacorogna. *Direct methods in the calculus of variations*, volume 78 of *Applied Mathematical Sciences*. Springer, New York, second edition, 2008.
- [6] 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.
- [7] M. Grasmair. Generalized Bregman distances and convergence rates for non-convex regularization methods. *Inverse Probl.*, 26(11):115014, October 2010.
- [8] 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.
- [9] J. A. Iglesias, M. Rumpf, and O. Scherzer. Shape-aware matching of implicit surfaces based on thin shell energies. *Found. Comput. Math.*, Online First:1–37, 2017.

- [10] C. Kirisits and O. Scherzer. Convergence rates for regularization functionals with polyconvex integrands. *Inverse Probl.*, 33(8):085008, August 2017.
- [11] 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.
- [12] 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.
- [13] K. Yosida. *Functional Analysis*, volume 123 of *Die Grundlehren der Mathematischen Wissenschaften*. Academic Press Inc., New York, 1965.
- [14] C. A. Zarzer. On Tikhonov regularization with non-convex sparsity constraints. *Inverse Probl.*, 25:025006, 2009.