CONVEX REGULARIZATION OF DISCRETE-VALUED INVERSE PROBLEMS

Christian Clason[*]   Thi Bich Tram Do[*]

July 26, 2017

Abstract   This work is concerned with linear inverse problems where a distributed parameter is known *a priori* to only take on values from a given discrete set. This property can be promoted in Tikhonov regularization with the aid of a suitable convex but nondifferentiable regularization term. This allows applying standard approaches to show well-posedness and convergence rates in Bregman distance. Using the specific properties of the regularization term, it can be shown that convergence (albeit without rates) actually holds pointwise. Furthermore, the resulting Tikhonov functional can be minimized efficiently using a semi-smooth Newton method. Numerical examples illustrate the properties of the regularization term and the numerical solution.

1 INTRODUCTION

We consider Tikhonov regularization of inverse problems, where the unknown parameter to be reconstructed is a distributed function that only takes on values from a given discrete set (i.e., the values are known, but not in which points they are attained). Such problems can occur, e.g., in nondestructive testing or medical imaging; a similar task also arises as a sub-step in segmentation or labelling problems in image processing. The question we wish to address here is the following: If such strong *a priori* knowledge is available, how can it be incorporated in an efficient manner? Specifically, if X and Y are function spaces, $F : X \rightarrow Y$ denotes the parameter-to-observation mapping, and $y^\delta \in Y$ is the given noisy data, we would wish to solve the constrained Tikhonov functional

$$(1.1) \qquad \min_{u \in U} \frac{1}{2} \|F(u) - y^\delta\|_Y$$

for

$$(1.2) \qquad U := \{u \in X : u \in \{u_1, \ldots, u_d\} \text{ pointwise}\},$$

where $u_1, \ldots, u_d \in \mathbb{R}$ are the known parameter values. However, this set is nonconvex, and hence the functional in (1.1) is not weakly lower-semicontinuous and can therefore not be treated

---

[*]University of Duisburg-Essen, Faculty of Mathematics, Thea-Leymann-Str. 9, 45127 Essen, Germany (christian.clason@uni-due.de, tram.do@uni-due.de)

by standard techniques. (In particular, it will in general not admit a minimizer.) A common strategy to deal with such problems is by convex relaxation, i.e., replacing U by its convex hull

$$\operatorname{co} U = \{u \in X : u \in [u_1, u_d] \text{ pointwise}\} .$$

This turns (1.1) into a classical *bang-bang* problem, whose solution is known to generically take on only the values u_1 or u_d; see, e.g., [4, 24]. If $d > 2$, intermediate parameter values are therefore lost in the reconstruction. (Here we would like to remark that a practical regularization should not only converge as the noise level tends to zero but also yield informative reconstructions for fixed – and ideally, a large range of – noise levels.) As a remedy, we propose to add a convex regularization term that promotes reconstructions in U (rather than merely in $\operatorname{co} U$) for the convex relaxation. Specifically, we choose the convex integral functional

$$\mathcal{G} : X \to \mathbb{R}, \qquad \mathcal{G}(u) := \int g(u(x)) \, dx,$$

for a convex integrand $g : \mathbb{R} \to \mathbb{R}$ with a polyhedral epigraph whose vertices correspond to the known parameter values $u_1, \ldots, u_d$. Just as in L^1 regularization for sparsity (and in linear optimization), it can be expected that minimizers are found at the vertices, thus yielding the desired structure.

This approach was first introduced in [9] in the context of linear optimal control problems for partial differential equations, where the so-called *multi-bang* (as a generalization of bang-bang) penalty $\mathcal{G}$ was obtained as the convex envelope of a (nonconvex) L^0 penalization of the constraint $u \in U$. The application to nonlinear control problems and the limit as the L^0 penalty parameter tends to infinity were considered in [10], and our particular choice of $\mathcal{G}$ is based on this work. The extension of this approach to vector-valued control problems was carried out in [11].

Our goal here is therefore to investigate the use of the multi-bang penalty from [10] as a regularization term in inverse problems, in particular addressing convergence and convergence rates as the noise level and the regularization parameter tend to zero. Due to the convexity of the penalty, these follow from standard results on convex regularization if convergence is considered with respect to the Bregman distance. The main contribution of this work is to show that due to the structure of the pointwise penalty, this convergence can be shown to actually hold pointwise. Since the focus of our work is the novel convex regularization term, we restrict ourselves to linear problems for the sake of presentation. However, all results carry over in a straightforward fashion to nonlinear problems. Finally, we describe following [9, 10] the computation of Tikhonov minimizers using a path-following semismooth Newton method.

Let us briefly mention other related literature. Regularization with convex nonsmooth functionals is now a widely studied problem, and we only refer to the monographs [17, 21, 23] as well as the seminal works [6, 13, 15, 20]. To the best of our knowledge, this is the first work treating regularization of general inverse problems with discrete-valued distributed parameters. As mentioned above, similar problems occur frequently in image segmentation or, more generally, image labelling problems. The former are usually treated by (multi-phase) level set methods [27] or by a combination of total variation minimization and thresholding [7]. More general

approaches to image labelling problems are based on graph-cut algorithms [1, 16] or, more recently, vector-valued convex relaxation [14, 19]. Both multi-phase level sets and vector-valued relaxations, however, have the disadvantage that the dimension of the parameter space grows quickly with the number of admissible values, which is not the case in our approach. On the other hand, our approach assumes, similar to [16], a linear ordering of the desired values which is not necessary in the vector-valued case; see also [11].

This work is organized as follows. In Section 2, we give the concrete form of the pointwise multi-bang penalty g and summarize its relevant properties. Section 3 is concerned with well-posedness, convergence, and convergence rates of the corresponding Tikhonov regularization. Our main result, the pointwise convergence of the regularized solutions to the true parameter, is the subject of Section 4. We also briefly discuss the structure of minimizers for given y^δ and fixed $\alpha > 0$ in Section 5. Finally, we address the numerical solution of the Tikhonov minimization problem using a semismooth Newton method in Section 6 and apply this approach to an inverse source problem for a Poisson equation in Section 7.

2 MULTI-BANG PENALTY

Let $u_1 < \cdots < u_d \in \mathbb{R}$, $d \geq 2$, be the given admissible parameter values and $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N}$, be a bounded domain. Following [10, § 3], we define the corresponding multi-bang penalty

$$\mathcal{G} : L^2(\Omega) \to \overline{\mathbb{R}}, \qquad \mathcal{G}(u) = \int_\Omega g(u(x))\, dx,$$

for $g : \mathbb{R} \to \overline{\mathbb{R}}$ defined by

$$g(v) = \begin{cases} \frac{1}{2}((u_i + u_{i+1})v - u_i u_{i+1}) & \text{if } v \in [u_i, u_{i+1}], \quad 1 \leq i < d, \\ \infty & \text{else.} \end{cases}$$

(Note that we have now included the convex constraint $u \in \mathrm{co}\, U$ in the definition of $\mathcal{G}$.) This choice can be motivated as the convex hull of $\frac{1}{2}\|\cdot\|^2_{L^2(\Omega)} + \delta_U$, where δ_U denotes the indicator function of the set U defined in (1.2) in the sense of convex analysis, i.e., $\delta_U(u) = 0$ if $u \in U$ and ∞ else; see [10, § 3]. Setting

$$g_i(v) := \frac{1}{2}((u_i + u_{i+1})v - u_i u_{i+1}), \qquad 1 \leq i < d,$$

it is straightforward to verify that

$$g(v) = \max_{1 \leq i < d} g_i(v), \qquad v \in [u_1, u_d],$$

and hence g is the pointwise supremum of affine functions and therefore convex and continuous on the interior of its effective domain $\mathrm{dom}\, g = [u_1, u_d]$.

We can thus apply the sum rule and maximum rule of convex analysis (see, e.g., [22, Props. 4.5.1 and 4.5.2, respectively]), and obtain for the convex subdifferential at $v \in \operatorname{dom} g$ that

$$
\begin{aligned}
\partial g(v) &= \partial \left(\max_{1 \le i < d} g_i + \delta_{[u_1, u_d]} \right)(v) \\
&= \partial \left(\max_{1 \le i < d} g_i \right)(v) + \partial \delta_{[u_1, u_d]}(v) \\
&= \operatorname{co} \left(\bigcup_{i : g(v) = g_i(v)} g_i'(v) \right) + \partial \delta_{[u_1, u_d]}(v).
\end{aligned}
$$

Using the definition of g_i together with the classical characterization of the subdifferential of an indicator function via its normal cone yields the explicit characterization

(2.1)
$$
\partial g(v) = \begin{cases}
\left(-\infty, \frac{1}{2}(u_1 + u_2) \right] & \text{if } v = u_1, \\
\left\{ \frac{1}{2}(u_i + u_{i+1}) \right\} & \text{if } v \in (u_i, u_{i+1}), \quad 1 \le i < d, \\
\left[\frac{1}{2}(u_{i-1} + u_i), \frac{1}{2}(u_i + u_{i+1}) \right] & \text{if } v = u_i, \quad 1 < i < d, \\
\left[\frac{1}{2}(u_{d-1} + u_d), \infty \right) & \text{if } v = u_d, \\
\emptyset & \text{else.}
\end{cases}
$$

In Sections 5 and 6, we will also make use of the subdifferential of the Fenchel conjugate g^* of g. Here we can use the fact that g is convex and hence $q \in \partial g(v)$ if and only if $v \in \partial g^*(q)$ (see, e.g., [22, Prop. 4.4.4]) to obtain

(2.2)
$$
\partial g^*(q) \in \begin{cases}
\{u_1\} & \text{if } q \in \left(-\infty, \frac{1}{2}(u_1 + u_2) \right), \\
[u_i, u_{i+1}] & \text{if } q = \frac{1}{2}(u_i + u_{i+1}), \quad 1 \le i < d, \\
\{u_i\} & \text{if } q \in \left(\frac{1}{2}(u_{i-1} + u_i), \frac{1}{2}(u_i + u_{i+1}) \right), \quad 1 < i < d, \\
\{u_d\} & \text{if } q \in \left(\frac{1}{2}(u_{d-1} + u_d), \infty \right), \\
\emptyset & \text{else.}
\end{cases}
$$

(Note that subdifferentials are always closed.) We illustrate these characterizations for a simple example in Figure 1.

Finally, since g is proper, convex, and lower semi-continuous by construction, the corresponding integral functional $\mathcal{G} : L^2(\Omega) \to \overline{\mathbb{R}}$ is proper, convex and weakly lower semicontinuous as well; see, e.g., [2, Proposition 2.53]. Furthermore, the subdifferential can be computed pointwise as

(2.3)
$$
\partial \mathcal{G}(u) = \left\{ v \in L^2(\Omega) : v(x) \in \partial g(u(x)) \quad \text{for almost every } x \in \Omega \right\},
$$

see, e.g., [2, Prop. 2.53]. The same is true for the Fenchel conjugate $\mathcal{G}^* : L^2(\Omega) \to \overline{\mathbb{R}}$ and hence for $\partial \mathcal{G}^*$ (which is thus an element of $L^\infty(\Omega)$ instead of $L^2(\Omega)$); see, e.g., [12, Props. IV.1.2, IX.2.1].
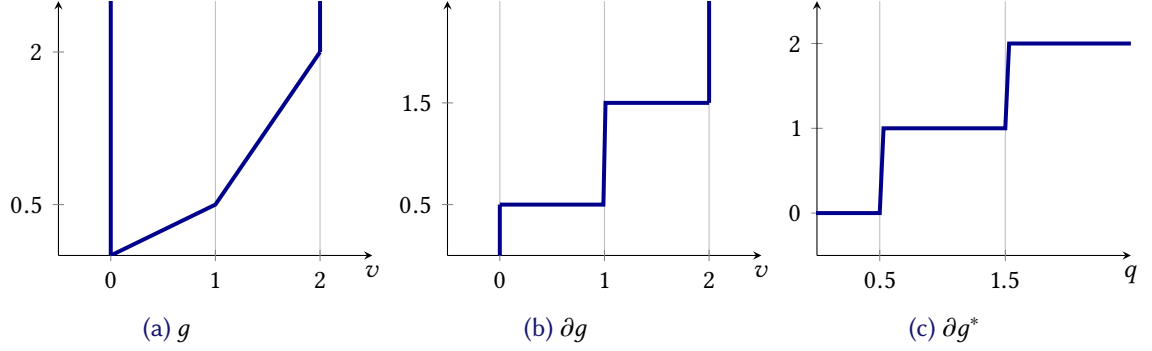
4

Figure 1: Structure of pointwise multibang penalty for the choice $(u_1, u_2, u_3) = (0, 1, 2)$

3 MULTI-BANG REGULARIZATION

We consider for a linear operator $K : X \to Y$ between the Hilbert spaces $X = L^2(\Omega)$ and Y and exact data $y^\dagger \in Y$ the inverse problem of finding $u \in X$ such that

(3.1) $$Ku = y^\dagger.$$

We assume that K is weakly closed, i.e., $u_n \rightharpoonup u$ and $Ku_n \rightharpoonup y$ imply $y = Ku$. For the sake of presentation, we also assume that (3.1) admits a solution $u^\dagger \in X$. Let now $y^\delta \in Y$ be given noisy data with $\|y^\delta - y^\dagger\|_Y \le \delta$ for some noise level $\delta > 0$. The *multi-bang regularization* of (3.1) for $\alpha > 0$ then consists in solving

(3.2) $$\min_{u \in X} \frac{1}{2}\|Ku - y^\delta\|_Y^2 + \alpha \mathcal{G}(u).$$

Since $\mathcal{G}$ is proper, convex and semi-continuous with bounded effective domain $\mathrm{co}\, U$, and K is weakly closed, the following results can be proved by standard semi-continuity methods; see also [10, 11].

Proposition 3.1 (Existence and uniqueness). *For every $\alpha > 0$, there exists a minimizer u_α^δ to (3.2). If K is injective, this minimizer is unique.*

Proposition 3.2 (Stability). *Let $\{y_n\}_{n \in \mathbb{N}} \subset Y$ be a sequence converging strongly to $y^\delta \in Y$ and $\alpha > 0$ be fixed. Then the corresponding sequence of minimizers $\{u_n\}_{n \in \mathbb{N}}$ to (3.2) contains a subsequence converging weakly to a minimizer u_α^δ.*

We now address convergence for $\delta \to 0$. Recall that an element $u^\dagger \in X$ is called a $\mathcal{G}$-minimizing solution to (3.1) if it is a solution to (3.1) and $\mathcal{G}(u^\dagger) \le \mathcal{G}(u)$ for all solutions u to (3.1). The following result is standard as well; see, e.g., [17, 21, 23].

Proposition 3.3 (Convergence). *Let $\{y^{\delta_n}\}_{n \in \mathbb{N}} \subset Y$ be a sequence of noisy data with $\|y^{\delta_n} - y^\dagger\|_Y \le \delta_n \to 0$, and choose $\alpha_n := \alpha_n(\delta_n)$ satisfying*

$$\lim_{n \to \infty} \frac{\delta_n^2}{\alpha_n} = 0 \qquad and \qquad \lim_{n \to \infty} \alpha_n = 0.$$

Then the corresponding sequence of minimizers $\{u_{\alpha_n}^{\delta_n}\}_{n\in\mathbb{N}}$ to (3.2) contains a subsequence converging weakly to a $\mathcal{G}$-minimizing solution $u^\dagger$.

For convex nonsmooth regularization terms, convergence rates are usually derived in terms of the Bregman distance [5], which is defined for $u_1, u_2 \in X$ and $p_1 \in \partial\mathcal{G}(u_1)$ as

$$d_{\mathcal{G}}^{p_1}(u_2, u_1) = \mathcal{G}(u_2) - \mathcal{G}(u_1) - \langle p_1, u_2 - u_1 \rangle_X.$$

From the convexity of $\mathcal{G}$, it follows that $d_{\mathcal{G}}^{p_1}(u_2, u_1) \geq 0$ for all $u_2 \in X$. Furthermore, we have from, e.g., [17, Lem. 3.8] the so-called *three-point identity*

$$(3.3) \qquad d_{\mathcal{G}}^{p_1}(u_3, u_1) = d_{\mathcal{G}}^{p_2}(u_3, u_2) + d_{\mathcal{G}}^{p_1}(u_2, u_1) + (p_2 - p_1)(u_3 - u_2)$$

for any $u_1, u_2, u_3 \in X$ and $p_1 \in \mathcal{G}(u_1)$ and $p_2 \in \partial\mathcal{G}(u_2)$. Finally, we point out that due to the pointwise characterization (2.3) of the subdifferential of the integral functional $\mathcal{G}$, we have that

$$(3.4) \qquad d_{\mathcal{G}}^p(u_2, u_1) = \int_\Omega d_g^{p(x)}(u_2(x), u_1(x))dx$$

for

$$d_g^q(v_2, v_1) = g(v_2) - g(v_1) - q(v_2 - v_1).$$

Standard arguments can then be used to show convergence rates for *a priori* and *a posteriori* parameter choice rules under the usual source conditions; see, e.g., [6, 17, 20, 21, 23]. Here we follow the latter and assume that there exists a $w \in Y$ such that

$$(3.5) \qquad p^\dagger := K^* w \in \partial\mathcal{G}(u^\dagger).$$

Under the *a priori* choice rule

$$(3.6) \qquad \alpha = c\delta \qquad \text{for some } c > 0,$$

we obtain the following convergence rate from, e.g., [17, Cor. 3.4].

Proposition 3.4 (Convergence rate, *a priori*). *Assume that the source condition* (3.5) *holds and that* $\alpha = \alpha(\delta)$ *is chosen according to* (3.6). *Then there exists a* $C > 0$ *such that*

$$d_{\mathcal{G}}^{p^\dagger}(u_\alpha^\delta, u^\dagger) \leq C\delta.$$

We obtain the same rate under the classical Morozov discrepancy principle

$$(3.7) \qquad \delta < \|Ku_\alpha^\delta - y^\delta\|_Y \leq \tau\delta,$$

for some $\tau > 1$ from, e.g., [17, Thm. 3.15].

Proposition 3.5 (Convergence rate, *a posteriori*). *Assume that the source condition* (3.5) *holds and that* $\alpha = \alpha(\delta)$ *is chosen according to* (3.7). *Then there exists a* $C > 0$ *such that*

$$d_{\mathcal{G}}^{p^\dagger}(u_\alpha^\delta, u^\dagger) \leq C\delta.$$

4 POINTWISE CONVERGENCE

The pointwise definition (3.4) of the Bregman distance together with the explicit pointwise characterization (2.1) of subgradients allows us to show that the convergence in Proposition 3.3 is actually pointwise if $u^\dagger(x) \in \{u_1, \dots, u_d\}$ almost everywhere. The following lemma provides the central argument for pointwise convergence.

Lemma 4.1. *Let $v^\dagger \in \{u_1, \dots, u_d\}$ and $q^\dagger \in \partial g(v^\dagger)$ satisfying*

$$
(4.1) \qquad q^\dagger \in \begin{cases} \left\{ \frac{1}{2}(u_i + u_{i+1}) \right\} & \text{if } v^\dagger \in (u_i, u_{i+1}), \quad 1 \le i < d, \\ \left(\frac{1}{2}(u_i + u_{i-1}), \frac{1}{2}(u_i + u_{i+1}) \right), & \text{if } v^\dagger = u_i, \qquad\quad 1 < i < d \\ \left(-\infty, \frac{1}{2}(u_1 + u_2) \right), & \text{if } v^\dagger = u_1 \\ \left(\frac{1}{2}(u_d + u_{d-1}), \infty \right), & \text{if } v^\dagger = u_d \end{cases}
$$

Furthermore, let $\{v_n\}_{n \in \mathbb{N}} \subset [u_1, u_d]$ be a sequence with

$$
d_g^{q^\dagger}(v_n, v^\dagger) \to 0.
$$

Then, $v_n \to v^\dagger$.

Proof. We argue by contraposition: Assume that v_n does not converge to $v^\dagger = u_i$ for some $1 \le i \le d$. Then there exists an $\varepsilon > 0$ such that for every $n_0 \in \mathbb{N}$, there is an $n \ge n_0$ with $|v_n - v^\dagger| > \varepsilon$, i.e., either $v_n > u_i + \varepsilon$ or $v_n < u_i - \varepsilon$. We now further discriminate these two cases. (Note that some cases cannot occur if $i = 1$ or $i = d$.)

(i) $v_n > u_{i+1}$: Then, $v_n \in (u_k, u_{k+1}]$ for some $k \ge i + 1$. The three point identity (3.3) yields that
$$
d_g^{q^\dagger}(v_n, v^\dagger) = d_g^{q_{i+1}}(v_n, u_{i+1}) + d_g^{q^\dagger}(u_{i+1}, v^\dagger) + (q_{i+1} - q^\dagger)(v_n - u_{i+1})
$$
for $q_{i+1} \in \partial g(u_{i+1})$. We now estimate each term separately. The first term is nonnegative by the properties of Bregman distances. For the last term, we can use the assumption (4.1) and the pointwise characterization (2.1) to obtain

$$
q^\dagger \in \left(\tfrac{1}{2}(u_i + u_{i-1}), \tfrac{1}{2}(u_i + u_{i+1}) \right) \quad \text{and} \quad q_{i+1} \in \left[\tfrac{1}{2}(u_{i+1} + u_i), \tfrac{1}{2}(u_{i+1} + u_{i+2}) \right],
$$

which implies that $q_{i+1} - q^\dagger > 0$. By assumption we have $v_n - u_{i+1} > 0$, which together implies that the last term is strictly positive. For the second term, we can use that $v^\dagger, u_{i+1} \in [u_i, u_{i+1}]$ to simplify the Bregman distance to

$$
d_g^{q^\dagger}(u_{i+1}, v^\dagger) = \frac{1}{2}(u_{i+1} - u_i)(u_{i+1} + u_i - 2q^\dagger) > 0,
$$

again by assumption (4.1). Since this term is independent of n, we obtain the estimate

$$
d_g^{q^\dagger}(v_n, v^\dagger) > d_g^{q^\dagger}(u_{i+1}, v^\dagger) =: \varepsilon_1 > 0.
$$

(ii) $u_i < v_n \le u_{i+1}$: In this case, we can again simplify

$$d_g^{q^\dagger}(v_n, v^\dagger) = \frac{1}{2}(u_{i+1} + u_i - 2q^\dagger)(v_n - v^\dagger) > C_1\varepsilon,$$

since $C_1 := \frac{1}{2}(u_{i+1} + u_i - 2q^\dagger) > 0$ by assumption (4.1) and $v_n - v^\dagger > \varepsilon$ by hypothesis.

(iii) $v_n < u_i$: We argue similarly to either obtain

$$d_g^{q^\dagger}(v_n, v^\dagger) > d_g^{q^\dagger}(u_{i-1}, v^\dagger) =: \varepsilon_2 > 0$$

or

$$d_g^{q^\dagger}(v_n, v^\dagger) > C_2\varepsilon$$

for $C_2 := -\frac{1}{2}(u_{i-1} + u_i - 2q^\dagger) > 0$.

Thus if we set $\tilde{\varepsilon} := \min\{\varepsilon_1, \varepsilon_2, C_1\varepsilon, C_2\varepsilon\}$, for every $n_0 \in \mathbb{N}$ we can find $n \ge n_0$ such that $d_g^{q^\dagger}(v_n, v^\dagger) > \tilde{\varepsilon} > 0$. Hence, $d_g^{q^\dagger}(v_n, v^\dagger)$ cannot converge to 0. □

Assumption (4.1) can be interpreted as a strict complementarity condition for $q^\dagger$ and $v^\dagger$. Comparing (4.1) to (2.1), we point out that such a choice of $q^\dagger$ is always possible. If $v^\dagger \notin \{u_1, \ldots, u_d\}$, on the other hand, convergence in Bregman distance is uninformative.

Lemma 4.2. *Let $v^\dagger \in (u_i, u_{i+1})$ for some $1 \le i < d$ and $q^\dagger \in \partial g(v^\dagger)$. Then we have*

$$d_{\mathcal{G}}^{q^\dagger}(v, v^\dagger) = 0 \qquad \text{for any} \quad v \in [u_i, u_{i+1}].$$

Proof. By the definition of the Bregman distance and the characterization (2.1) of $\partial g(v^\dagger)$ (which is single-valued under the assumption on $v^\dagger$), we directly obtain

$$d_g^{q^\dagger}(v, v^\dagger) = \frac{1}{2}\left[(u_i + u_{i+1})v - u_i u_{i+1}\right] - \frac{1}{2}\left[(u_i + u_{i+1})v^\dagger - u_i u_{i+1}\right]$$
$$- \frac{1}{2}(u_i + u_{i+1})(v - v^\dagger) = 0$$

for any $v \in [u_i, u_{i+1}]$. □

Lemma 4.1 allows us to translate the weak convergence from Proposition 3.3 to pointwise convergence, which is the main result of our work.

Theorem 4.3. *Assume the conditions of Proposition 3.3 hold. If $u^\dagger(x) \in \{u_1, \ldots, u_d\}$ almost everywhere, the subsequence $u_{\alpha_n}^{\delta_n} \to u^\dagger$ pointwise almost everywhere.*

Proof. From Proposition 3.3, we obtain a subsequence $\{u_n\}_{n \in \mathbb{N}}$ of $\{u_{\alpha_n}^{\delta_n}\}_{n \in \mathbb{N}}$ converging weakly to $u^\dagger$. Since $\mathcal{G}$ is convex and lower semicontinuous, we have that

(4.2) $$\mathcal{G}(u^\dagger) \le \liminf_{n \to \infty} \mathcal{G}(u_n) \le \lim_{n \to \infty} \mathcal{G}(u_n).$$

By the minimizing properties of $\{u_n\}_{n \in \mathbb{N}}$ and the nonnegativity of the discrepancy term, we further obtain that

$$\alpha_n \mathcal{G}(u_n) \leq \frac{1}{2}\|Ku_n - y^{\delta_n}\|_Y^2 + \alpha_n \mathcal{G}(u_n) \leq \frac{\delta_n^2}{2} + \alpha_n \mathcal{G}(u^\dagger).$$

Dividing this inequality by α_n and passing to the limit $n \to \infty$, the assumption on α_n from Proposition 3.3 yields that

$$\lim_{n \to \infty} \mathcal{G}(u_n) \leq \mathcal{G}(u^\dagger),$$

which combined with (4.2) gives $\lim_{n \to \infty} \mathcal{G}(u_n) = \mathcal{G}(u^\dagger)$. Hence, $u_n \rightharpoonup u^\dagger$ implies that $d_{\mathcal{G}}^{p^\dagger}(u_n, u^\dagger) \to 0$ for any $p^\dagger \in \partial \mathcal{G}(u^\dagger)$. By the pointwise characterization (3.4) and the nonnegativity of Bregman distances, this implies that $d_g^{p^\dagger(x)}(u_n(x), u^\dagger(x)) \to 0$ for almost every $x \in \Omega$. Choosing now $p^\dagger \in \partial \mathcal{G}(u^\dagger)$ such that (4.1) holds for $q^\dagger = p^\dagger(x)$ and $v^\dagger = u^\dagger(x)$ almost everywhere, the claim follows from Lemma 4.1. $\qquad\square$

Since $u_n(x) \in [u_1, u_d]$ by construction, the subsequence $\{u_n\}_{n \in \mathbb{N}}$ is bounded in $L^\infty(\Omega)$ and hence also converges strongly in $L^p(\Omega)$ for any $1 \leq p < \infty$ by Lebesgue's dominated convergence theorem. We remark that since Lemma 4.1 applied to $u_n(x)$ and $u^\dagger(x)$ does not hold uniformly in Ω, we cannot expect that the convergence rates from Propositions 3.4 and 3.5 hold pointwise or strongly as well.

5 STRUCTURE OF MINIMIZERS

We now briefly discuss the structure of reconstructions obtained by minimizing the Tikhonov functional in (3.2) for given $y^\delta \in Y$ and fixed $\alpha > 0$, based on the necessary optimality conditions for (3.2). Since the discrepancy term is convex and differentiable, we can apply the sum rule for convex subdifferentials. Furthermore, the standard calculus for Fenchel conjugates and subdifferentials (see, e.g., [22]) yields for $\mathcal{G}_\alpha := \alpha \mathcal{G}$ that $\mathcal{G}_\alpha^*(p) = \alpha \mathcal{G}^*(\alpha^{-1}p)$ and hence that $p \in \partial \mathcal{G}_\alpha(u)$ if and only if $u \in \partial \mathcal{G}_\alpha^*(p) = \partial \mathcal{G}^*(\frac{1}{\alpha}p)$. We thus obtain as in [9] that $\bar{u} := u_\alpha^\delta \in L^2(\Omega)$ is a solution to (3.2) if and only if there exists a $\bar{p} \in L^2(\Omega)$ satisfying

(5.1)
$$\begin{cases} \bar{p} = K^*(y^\delta - K\bar{u}) \\ \bar{u} \in \partial \mathcal{G}_\alpha^*(\bar{p}) := \begin{cases} \{u_i\} & \bar{p}(x) \in Q_i, & 1 \leq i \leq d, \\ [u_i, u_{i+1}] & \bar{p}(x) \in Q_{i,i+1} & 1 \leq i < d. \end{cases} \end{cases}$$

for

$$\begin{aligned} Q_1 &= \left\{ q : q < \tfrac{\alpha}{2}(u_1 + u_2) \right\}, \\ Q_i &= \left\{ q : \tfrac{\alpha}{2}(u_{i-1} + u_i) < q < \tfrac{\alpha}{2}(u_i + u_{i+1}) \right\}, \quad 1 < i < d, \\ Q_d &= \left\{ q : q > \tfrac{\alpha}{2}(u_{d-1} + u_d) \right\}, \\ Q_{i,i+i} &= \left\{ q : q = \tfrac{\alpha}{2}(u_i + u_{i+1}) \right\}, \qquad\qquad 1 \leq i < d. \end{aligned}$$

Here we have made use of the pointwise characterization in (2.2) and reformulated the case distinction in terms of $\bar{p}(x)$ instead of $\frac{1}{\alpha}\bar{p}(x)$.

First, we obtain directly from (5.1) the desired structure of the reconstruction $\bar{u}$: Apart from a singular set
$$\mathcal{S} := \left\{ x \in \Omega : \bar{p}(x) = \tfrac{\alpha}{2}(u_i + u_{i+1}) \text{ for some } 1 \le i < d \right\},$$
we always have $\bar{u}(x) \in \{u_1, \ldots, u_d\}$. For operators K where $K^* w$ cannot be constant on a set of positive measure unless $w = 0$ locally (as is the case for many operators involving solutions to partial differential equations; see [9, Prop. 2.3]) and $y^\delta \notin \operatorname{ran} K$, the singular set $\mathcal{S}$ has zero measure and hence the "multi-bang" structure $\bar{u} \in \{u_1, \ldots, u_d\}$ almost everywhere can be guaranteed *a priori* for any $\alpha > 0$.

Furthermore, we point out that the regularization parameter α only enters via the case distinction. In particular, increasing α shifts the conditions on $\bar{u}(x)$ such that the smaller values among the u_i become more preferred. In fact, if $\bar{p}$ is bounded, we can expect that there exists an $\alpha_0 > 0$ such that $\bar{u} \equiv u_1$ for all $\alpha > \alpha_0$. Conversely, for $\alpha \to 0$, the second line of (5.1) reduces to

$$\bar{u}(x) \in \begin{cases} \{u_1\} & \text{if } \bar{p}(x) < 0, \\ \{u_d\} & \text{if } \bar{p}(x) > 0, \\ [u_1, u_d] & \text{if } \bar{p}(x) = 0, \end{cases}$$

i.e., (5.1) coincides with the well-known optimality conditions for bang-bang control problems; see, e.g., [25, Lem. 2.26]. Since in the context of inverse problems, we only have $\alpha = \alpha(\delta) \to 0$ if $\delta \to 0$, the limit system (5.1) will contain consistent data and hence $\bar{p} \equiv 0$. This allows recovery of $u^\dagger(x) \in \{u_2, \ldots, u_{d-1}\}$ on a set of positive measure, consistent with Proposition 3.3. However, if $u^\dagger(x) \in \{u_1, \ldots, u_d\}$ does not hold almost everywhere, we can only expect weak and not strong convergence, cf. [11, Prop. 5.10 (ii)].

6 NUMERICAL SOLUTION

In this section we address the numerical solution of the Tikhonov minimization problem (3.2) for given $y^\delta \in Y$ and $\alpha > 0$, following [10]. For the sake of presentation, we omit the dependence on α and δ from here on. We start from the necessary (and, due to convexity, sufficient) optimality conditions (5.1). To apply a semismooth Newton method, we replace the subdifferential inclusion $\bar{u} \in \partial \mathcal{G}_\alpha^*(\bar{p})$ by its single-valued Moreau–Yosida regularization, i.e., we consider for $\gamma > 0$ the regularized optimality conditions

(6.1)
$$\begin{cases} p_\gamma = K^*(y^\delta - K u_\gamma) \\ u_\gamma = (\partial \mathcal{G}_\alpha^*)_\gamma (p_\gamma). \end{cases}$$

The Moreau–Yosida regularization can also be expressed as

$$H_\gamma := (\partial \mathcal{G}_\alpha^*)_\gamma = \partial (\mathcal{G}_{\alpha,\gamma})^*$$

for

$$\mathcal{G}_{\alpha,\gamma}(u) := \alpha \mathcal{G}(u) + \frac{\gamma}{2} \|u\|_{L^2(\Omega)}^2,$$

see, e.g., [3, Props. 13.21, 12.29]. This implies that for (u_γ, p_γ) satisfying (6.1), u_γ is a solution to the strictly convex problem

$$\min_{u \in L^2(\Omega)} \frac{1}{2} \|Ku - y^\delta\|_Y^2 + \alpha \mathcal{G}(u) + \frac{\gamma}{2} \|u\|_{L^2(\Omega)}^2,$$

so that existence of a solution can be shown by the same arguments as for (3.2). Note that by regularizing the conjugate subdifferential, we have not smoothed the nondifferentiability but merely made the functional (more) strongly convex. The regularization of $\mathcal{G}_\alpha^*$ instead of $\mathcal{G}^*$ also ensures that the regularization is robust for $\alpha \to 0$. From [10, Prop. 4.1], we obtain the following convergence result.

Proposition 6.1. *The family $\{u_\gamma\}_{\gamma > 0}$ satisfying (6.1) contains at least one subsequence $\{u_{\gamma_n}\}_{n \in \mathbb{N}}$ converging to a global minimizer of (3.2) as $n \to \infty$. Furthermore, for any such subsequence, the convergence is strong.*

From [8, Appendix A.2] we further obtain the pointwise characterization

$$[H_\gamma(p)](x) = \begin{cases} u_i & \text{if } p(x) \in Q_i^\gamma, & 1 \le i \le d, \\ \frac{1}{\gamma}(p(x) - \frac{\alpha}{2}(u_i + u_{i+1})) & \text{if } p(x) \in Q_{i,i+1}^\gamma, & 1 \le i < d, \end{cases}$$

where

$$\begin{aligned} Q_1^\gamma &= \left\{ q : q < \frac{\alpha}{2}((1 + 2\gamma)u_1 + u_2) \right\}, \\ Q_i^\gamma &= \left\{ q : \frac{\alpha}{2}(u_{i-1} + (1 + 2\gamma)u_i) < q < \frac{\alpha}{2}((1 + 2\gamma)u_i + u_{i+1}) \right\} \quad \text{for } 1 < i < d, \\ Q_d^\gamma &= \left\{ q : \frac{\alpha}{2}(u_{d-1} + (1 + 2\gamma)u_d) < q \right\}, \\ Q_{i,i+1}^\gamma &= \left\{ q : \frac{\alpha}{2}((1 + 2\gamma)u_i + u_{i+1}) \le q \le \frac{\alpha}{2}(u_i + (1 + 2\gamma)u_{i+1}) \right\} \quad \text{for } 1 \le i < d. \end{aligned}$$

Since H_γ is a superposition operator defined by a Lipschitz continuous and piecewise differentiable scalar function, H_γ is Newton-differentiable from $L^r(\Omega) \to L^2(\Omega)$ for any $r > 2$; see, e.g., [18, Example 8.12] or [26, Theorem 3.49]. A Newton derivative at p in direction h is given pointwise almost everywhere by

$$[D_N H_\gamma(p)h](x) = \begin{cases} \frac{1}{\gamma} h(x) & \text{if } p(x) \in Q_{i,i+1}^\gamma, & 1 \le i < d, \\ 0 & \text{else.} \end{cases}$$

Hence if the range of K^* embeds into $L^r(\Omega)$ for some $r > 2$ (which is the case, e.g., for many convolution operators and solution operators for partial differential equations) and the semismooth Newton step is uniformly invertible, the corresponding Newton iteration converges locally superlinearly. We address this for the concrete example considered in the next section. In practice, the local convergence can be addressed by embedding the Newton method into a continuation strategy, i.e., starting for γ large and then iteratively reducing γ, using the previous solution as a starting point.

7 NUMERICAL EXAMPLES

We illustrate the proposed approach for an inverse source problem for the Poisson equation, i.e., we choose $K = A^{-1} : L^2(\Omega) \to L^2(\Omega)$ for $\Omega = [0,1]^2$ and $A = -\Delta$ together with homogeneous boundary conditions. We note that since Ω is a Lipschitz domain, we have that $\operatorname{ran} A^{-*} = \operatorname{ran} A^{-1} = H^2(\Omega) \cap H_0^1(\Omega)$, and hence this operator satisfies the conditions discussed in Section 5 that guarantee that $u_\alpha^\delta(x) \in \{u_1, \dots, u_d\}$ almost everywhere if $y^\delta \notin \operatorname{ran} K$; see [9, Prop. 2.3]. For the computational results below, we use a finite element discretization on a uniform triangular grid with 256×256 vertices.

The specific form of K can be used to reformulate the optimality condition (and hence the Newton system) into a more convenient form. Introducing $y_\gamma = A^{-1} u_\gamma$ and eliminating u_γ using the second relation of (6.1), we obtain as in [9] the equivalent system

$$(7.1) \qquad \begin{cases} A^* p_\gamma + y_\gamma - y^\delta = 0, \\ \quad A y_\gamma - H_\gamma(p_\gamma) = 0. \end{cases}$$

Setting $V := H_0^1(\Omega)$, we can consider this as an equation from $V \times V$ to $V^* \times V^*$, which due to the embedding $V \hookrightarrow L^p(\Omega)$ for $p > 2$ provides the necessary norm gap for Newton differentiability of H_γ. By the chain rule for Newton derivatives from, e.g., [18, Lem. 8.4], the corresponding Newton step therefore consists of solving for $(\delta y, \delta p) \in V \times V$ given $(y^k, p^k) \in V \times V$ in

$$(7.2) \qquad \begin{pmatrix} \mathrm{Id} & A^* \\ A & -D_N H_\gamma(p^k) \end{pmatrix} \begin{pmatrix} \delta y \\ \delta p \end{pmatrix} = - \begin{pmatrix} A^* p^k + y - y^\delta \\ A y^k - H_\gamma(p^k) \end{pmatrix}$$

and setting

$$y^{k+1} = y^k + \delta y, \qquad p^{k+1} = p^k + \delta p.$$

Note that the reformulated Newton matrix is symmetric, which in general is not the case for nonsmooth equations. Following [9, Prop. 4.3], the Newton step (7.2) is uniformly boundedly invertible, from which local superlinear convergence to a solution of (7.1) follows.

In practice, we include the continuation strategy described above as well as a simple backtracking line search based on the residual norm in (7.1) to improve robustness. Since the forward operator is linear and H_γ is piecewise linear, the semi-smooth Newton method has the following finite termination property: If $H_\gamma(p^{k+1}) = H_\gamma(p^k)$, then (y^{k+1}, p^{k+1}) satisfy (7.1); cf. [18, Rem. 7.1.1]. We then recover $u^{k+1} = H_\gamma(p^{k+1})$. In the implementation, we also terminate if more than 100 Newton iterations are performed, in which case the continuation is also terminated and the last successful iterate is returned. Otherwise we terminate if $\gamma < 10^{-12}$. In all results reported below, the continuation is terminated successfully. The implementation of this approach used to obtain the following results can be downloaded from https://github.com/clason/discreteregularization.

The first example illustrates the convergence behavior of the Tikhonov regularization. Here, the true parameter is chosen as

$$(7.3) \qquad u^\dagger(x) = u_1 + u_2\, \chi_{\{x : (x_1 - 0.45)^2 + (x_2 - 0.55)^2 < 0.1\}}(x)$$
$$+ (u_3 - u_2)\, \chi_{\{x : (x_1 - 0.4)^2 + (x_2 - 0.6)^2 < 0.02\}}(x)$$

for $(u_1, u_2, u_3) = (0, 0.1, 0.15)$; see Figure 2a. (This might correspond to, e.g., material properties of background, healthy tissue, and tumor, respectively.) The noisy data is constructed pointwise via

$$y^\delta = y^\dagger + (\tilde{\delta}\|y^\dagger\|_\infty)\xi,$$

where ξ is a vector of identically and independently normally distributed random variables with mean 0 and variance 1, and $\tilde{\delta} \in \{2^0, \ldots, 2^{-20}\}$. For each value of $\tilde{\delta}$, the corresponding regularization parameter α is chosen according to the discrepancy principle (3.7) with $\tau = 1.1$. Details on the convergence history are reported in Table 1, which shows the effective noise level $\delta := \|y^\delta - y^\dagger\|_2$, the parameter α selected as satisfying the Morozov discrepancy principle, the L^2-error $e_2 := \|u_\alpha^\delta - u^\dagger\|_2$ and the L^∞-error $e_\infty := \|u_\alpha^\delta - u^\dagger\|_\infty$. First, we note that the *a posteriori* choice approximately follows the *a priori* choice $\alpha \sim \delta$. Similarly, for larger values of δ, the L^2-error behaves as $e_2 \sim \delta$, which is no longer true for $\delta \to 0$ (and cannot be expected due to the nonsmooth regularization). The L^∞-error e_∞ is initially dominated by the jump in admissible parameter values: As long as there is a single point $x \in \Omega$ with $u_\alpha^\delta(x) = u_i \neq u_j = u^\dagger(x)$, we necessarily have $e_\infty \geq \min_{1 \leq i < d} u_{i+1} - u_i$. (Recall that we do not have a convergence rate and thus an error bound for pointwise convergence.) Later, e_∞ becomes smaller than this threshold value, which indicates that apart from points in the regularized singular set (i.e., where $p_\gamma(x) \in Q_{i,i+1}^\gamma$, which in these cases happens for 20 out of 256×256 vertices), the reconstruction is exact. Here we point out that since γ is independent of α, the Moreau–Yosida regularization for fixed γ becomes more and more active as $\alpha \to 0$. Nevertheless, in all cases $\gamma \ll \alpha$, and hence the multi-bang regularization dominates.

The pointwise convergence can also be seen clearly from Figure 2, which shows the true parameter $u^\dagger$ together with three representative reconstructions for different noise levels. It can be seen that for large noise, the corresponding large regularization suppresses the smaller inclusion; see Figure 2b. This is consistent with the discussion at the end of Section 5. For smaller noise, the inclusion is recovered well (Figure 2c), and for $\delta \approx 3.69 \cdot 10^{-4}$, the reconstruction is visually indistinguishable from the true parameter (Figure 2d).

The behavior is essentially the same if we set $(u_1, u_2, u_3) = (0, 0.1, 0.11)$ in (7.3) (i.e., a contrast of 10% instead of 50% for the inner inclusion), demonstrating the robustness of the multi-bang regularization; see Figure 3 and Table 2.

To illustrate the behavior if the true parameter does not satisfy the assumption $u^\dagger \in \{u_1, \ldots, u_d\}$ almost everywhere, we repeat the above for

$$u^\dagger(x) = u_1 + u_2\,\chi_{\{x:(x_1-0.45)^2+(x_2-0.55)^2<0.1\}}(x)$$
$$+ (u_3 - u_2)(1 - x_1)\,\chi_{\{x:(x_1-0.4)^2+(x_2-0.6)^2<0.02\}}(x)$$

with $(u_1, u_2, u_3) = (0, 0.1, 0.12)$; see Figure 4a. While for large noise level and regularization parameter value, the multi-bang regularization behaves as before (see Figure 4b), the reconstruction for smaller noise and regularization (Figure 4c) shows the typical checkerboard pattern expected from weak but not strong convergence; cf. [9, Rem. 4.2]. Nevertheless, as $\delta \to 0$, we still observe convergence to the true parameter; see Figure 4d and Table 3.

Finally, we address the qualitative dependence of the reconstruction on the regularization parameter α. Figure 5 shows reconstructions for the true parameter $u^\dagger$ from (7.3) again with $(u_1, u_2, u_3) = (0, 0.1, 0.15)$ for an effective noise level $\delta \approx 0.759$ and different values of α. First,
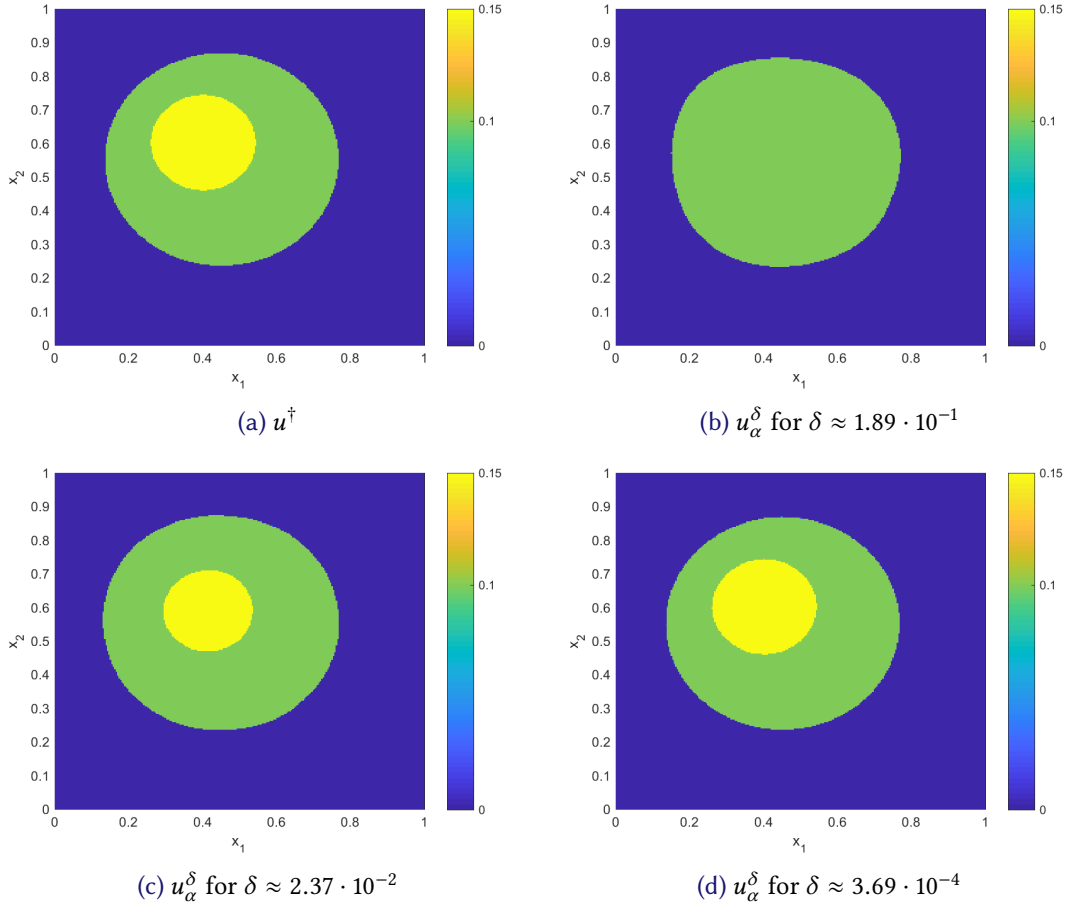
(a) $u^{\dagger}$

(b) u_{α}^{δ} for $\delta \approx 1.89 \cdot 10^{-1}$

(c) u_{α}^{δ} for $\delta \approx 2.37 \cdot 10^{-2}$

(d) u_{α}^{δ} for $\delta \approx 3.69 \cdot 10^{-4}$

Figure 2: True parameter $u^{\dagger}$ for $u_3 = 0.15$ and reconstructions u_{α}^{δ} for different values of δ

Table 1: Convergence behavior as $\delta \to 0$ for $u_3 = 0.15$: noise level δ, regularization parameter α, L^2-error e_2, L^{∞}-error e_{∞}

δ	α	e_2	e_{∞}	δ	α	e_2	e_{∞}
1.52	$1.00 \cdot 10^{-2}$	$1.60 \cdot 10^{1}$	$1.50 \cdot 10^{-1}$	$7.44 \cdot 10^{-4}$	$6.10 \cdot 10^{-7}$	$6.86 \cdot 10^{-1}$	$1.00 \cdot 10^{-1}$
$7.59 \cdot 10^{-1}$	$1.25 \cdot 10^{-3}$	8.64	$1.00 \cdot 10^{-1}$	$3.69 \cdot 10^{-4}$	$3.05 \cdot 10^{-7}$	$4.74 \cdot 10^{-1}$	$1.00 \cdot 10^{-1}$
$3.78 \cdot 10^{-1}$	$6.25 \cdot 10^{-4}$	6.18	$1.00 \cdot 10^{-1}$	$1.85 \cdot 10^{-4}$	$1.53 \cdot 10^{-7}$	$2.91 \cdot 10^{-1}$	$7.82 \cdot 10^{-2}$
$1.89 \cdot 10^{-1}$	$3.13 \cdot 10^{-4}$	4.26	$1.00 \cdot 10^{-1}$	$9.28 \cdot 10^{-5}$	$7.63 \cdot 10^{-8}$	$2.27 \cdot 10^{-1}$	$7.67 \cdot 10^{-2}$
$9.48 \cdot 10^{-2}$	$7.81 \cdot 10^{-5}$	4.32	$1.00 \cdot 10^{-1}$	$4.64 \cdot 10^{-5}$	$3.81 \cdot 10^{-8}$	$1.29 \cdot 10^{-1}$	$5.73 \cdot 10^{-2}$
$4.73 \cdot 10^{-2}$	$3.91 \cdot 10^{-5}$	3.67	$1.00 \cdot 10^{-1}$	$2.32 \cdot 10^{-5}$	$1.91 \cdot 10^{-8}$	$9.19 \cdot 10^{-2}$	$4.91 \cdot 10^{-2}$
$2.37 \cdot 10^{-2}$	$1.95 \cdot 10^{-5}$	2.97	$1.00 \cdot 10^{-1}$	$1.16 \cdot 10^{-5}$	$9.54 \cdot 10^{-9}$	$9.32 \cdot 10^{-2}$	$4.03 \cdot 10^{-2}$
$1.19 \cdot 10^{-2}$	$9.77 \cdot 10^{-6}$	2.33	$1.00 \cdot 10^{-1}$	$5.79 \cdot 10^{-6}$	$4.77 \cdot 10^{-9}$	$4.61 \cdot 10^{-2}$	$2.30 \cdot 10^{-2}$
$5.90 \cdot 10^{-3}$	$4.88 \cdot 10^{-6}$	1.76	$1.00 \cdot 10^{-1}$	$2.89 \cdot 10^{-6}$	$2.38 \cdot 10^{-9}$	$1.13 \cdot 10^{-1}$	$5.00 \cdot 10^{-2}$
$2.95 \cdot 10^{-3}$	$2.44 \cdot 10^{-6}$	1.33	$1.00 \cdot 10^{-1}$	$1.44 \cdot 10^{-6}$	$5.96 \cdot 10^{-10}$	$1.70 \cdot 10^{-2}$	$4.39 \cdot 10^{-3}$
$1.49 \cdot 10^{-3}$	$1.22 \cdot 10^{-6}$	$9.47 \cdot 10^{-1}$	$1.00 \cdot 10^{-1}$				

(a) $u^\dagger$

(b) u_α^δ for $\delta \approx 1.68 \cdot 10^{-1}$

(c) u_α^δ for $\delta \approx 2.17 \cdot 10^{-2}$

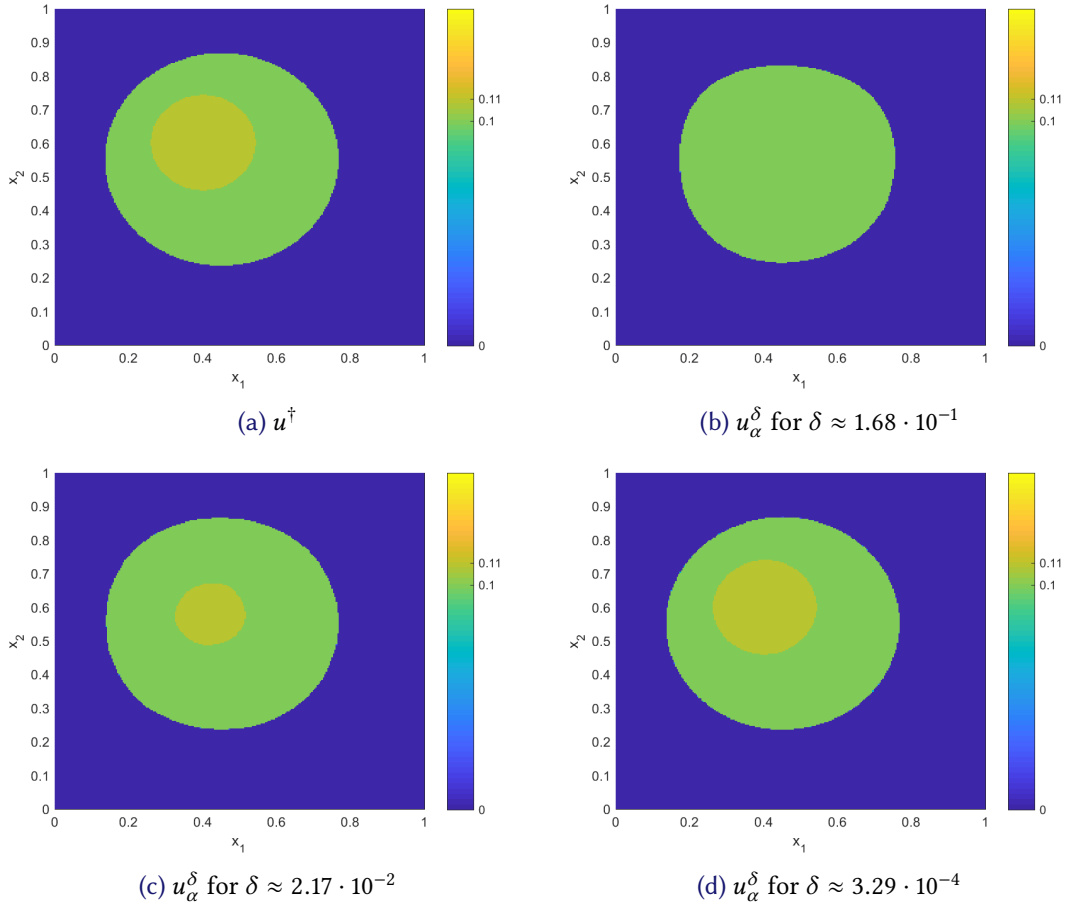(d) u_α^δ for $\delta \approx 3.29 \cdot 10^{-4}$

Figure 3: True parameter $u^\dagger$ for $u_3 = 0.11$ and reconstructions u_α^δ for different values of δ

Table 2: Convergence behavior as $\delta \to 0$ for $u_3 = 0.11$: noise level δ, regularization parameter α, L^2-error e_2, L^∞-error e_∞

δ	α	e_2	e_∞	δ	α	e_2	e_∞
1.34	$2.50 \cdot 10^{-3}$	1.16	$1.10 \cdot 10^{-1}$	$6.56 \cdot 10^{-4}$	$6.10 \cdot 10^{-7}$	$4.55 \cdot 10^{-1}$	$1.00 \cdot 10^{-1}$
$6.73 \cdot 10^{-1}$	$1.25 \cdot 10^{-3}$	9.13	$1.00 \cdot 10^{-1}$	$3.29 \cdot 10^{-4}$	$3.05 \cdot 10^{-7}$	$2.94 \cdot 10^{-1}$	$1.00 \cdot 10^{-1}$
$3.36 \cdot 10^{-1}$	$6.25 \cdot 10^{-4}$	6.89	$1.00 \cdot 10^{-1}$	$1.64 \cdot 10^{-4}$	$1.53 \cdot 10^{-7}$	$2.20 \cdot 10^{-1}$	$6.15 \cdot 10^{-2}$
$1.68 \cdot 10^{-1}$	$3.13 \cdot 10^{-4}$	4.91	$1.00 \cdot 10^{-1}$	$8.27 \cdot 10^{-5}$	$7.63 \cdot 10^{-8}$	$1.87 \cdot 10^{-1}$	$8.55 \cdot 10^{-2}$
$8.41 \cdot 10^{-2}$	$1.56 \cdot 10^{-4}$	3.27	$1.00 \cdot 10^{-1}$	$4.11 \cdot 10^{-5}$	$3.81 \cdot 10^{-8}$	$6.75 \cdot 10^{-2}$	$3.35 \cdot 10^{-2}$
$4.20 \cdot 10^{-2}$	$3.91 \cdot 10^{-5}$	1.90	$1.00 \cdot 10^{-1}$	$2.07 \cdot 10^{-5}$	$1.91 \cdot 10^{-8}$	$4.34 \cdot 10^{-2}$	$1.44 \cdot 10^{-2}$
$2.17 \cdot 10^{-2}$	$1.95 \cdot 10^{-5}$	1.57	$1.00 \cdot 10^{-1}$	$1.03 \cdot 10^{-5}$	$9.54 \cdot 10^{-9}$	$3.72 \cdot 10^{-2}$	$1.46 \cdot 10^{-2}$
$1.05 \cdot 10^{-3}$	$9.77 \cdot 10^{-6}$	1.19	$1.00 \cdot 10^{-1}$	$5.12 \cdot 10^{-6}$	$4.77 \cdot 10^{-9}$	$3.29 \cdot 10^{-2}$	$1.31 \cdot 10^{-2}$
$5.25 \cdot 10^{-3}$	$4.88 \cdot 10^{-6}$	$9.81 \cdot 10^{-1}$	$1.00 \cdot 10^{-1}$	$2.56 \cdot 10^{-6}$	$2.38 \cdot 10^{-9}$	$3.85 \cdot 10^{-2}$	$1.00 \cdot 10^{-2}$
$2.64 \cdot 10^{-3}$	$2.44 \cdot 10^{-6}$	$8.14 \cdot 10^{-1}$	$1.00 \cdot 10^{-1}$	$1.29 \cdot 10^{-6}$	$2.98 \cdot 10^{-10}$	$1.65 \cdot 10^{-1}$	$1.79 \cdot 10^{-2}$
$1.32 \cdot 10^{-4}$	$1.22 \cdot 10^{-6}$	$6.70 \cdot 10^{-1}$	$1.00 \cdot 10^{-1}$				

(a) $u^\dagger$

(b) u_α^δ for $\delta \approx 2.11 \cdot 10^{-2}$

(c) u_α^δ for $\delta \approx 3.29 \cdot 10^{-4}$

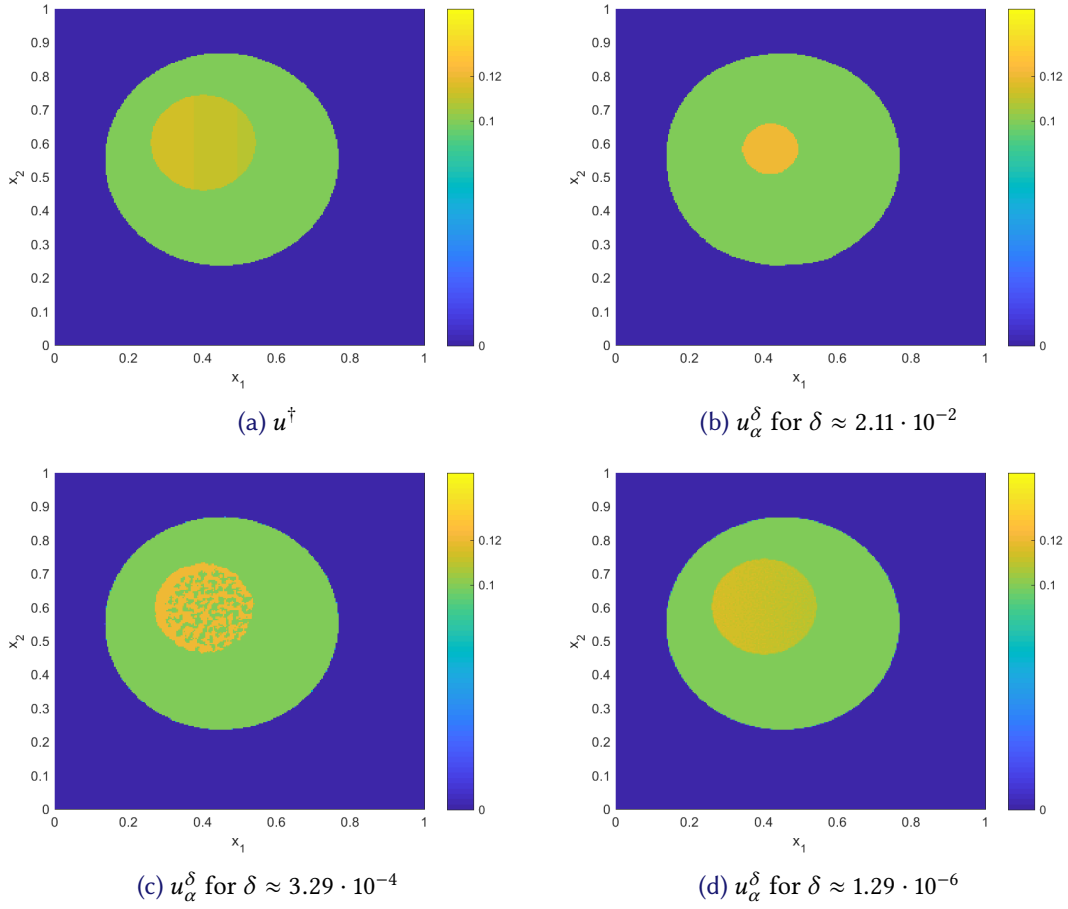(d) u_α^δ for $\delta \approx 1.29 \cdot 10^{-6}$

Figure 4: True parameter $u^\dagger$ and reconstructions u_α^δ for different values of δ

Table 3: Convergence behavior as $\delta \to 0$ for $u^\dagger$: noise level δ, regularization parameter α, L^2-error e_2, L^∞-error e_∞

δ	α	e_2	e_∞	δ	α	e_2	e_∞
1.36	$2.50 \cdot 10^{-3}$	$1.17 \cdot 10^1$	$1.15 \cdot 10^{-1}$	$6.60 \cdot 10^{-4}$	$6.10 \cdot 10^{-7}$	$8.46 \cdot 10^{-1}$	$1.00 \cdot 10^{-1}$
$6.77 \cdot 10^{-1}$	$1.25 \cdot 10^{-3}$	9.08	$1.00 \cdot 10^{-1}$	$3.29 \cdot 10^{-4}$	$1.53 \cdot 10^{-7}$	$7.23 \cdot 10^{-1}$	$1.00 \cdot 10^{-1}$
$3.39 \cdot 10^{-1}$	$6.25 \cdot 10^{-4}$	6.84	$1.00 \cdot 10^{-1}$	$1.66 \cdot 10^{-4}$	$7.63 \cdot 10^{-8}$	$6.20 \cdot 10^{-1}$	$5.63 \cdot 10^{-2}$
$1.69 \cdot 10^{-1}$	$3.12 \cdot 10^{-4}$	4.81	$1.00 \cdot 10^{-1}$	$8.25 \cdot 10^{-5}$	$3.81 \cdot 10^{-8}$	$6.04 \cdot 10^{-1}$	$5.60 \cdot 10^{-2}$
$8.48 \cdot 10^{-2}$	$1.56 \cdot 10^{-4}$	3.12	$1.00 \cdot 10^{-1}$	$4.12 \cdot 10^{-5}$	$1.91 \cdot 10^{-8}$	$5.69 \cdot 10^{-1}$	$1.83 \cdot 10^{-2}$
$4.22 \cdot 10^{-2}$	$3.91 \cdot 10^{-5}$	2.03	$1.00 \cdot 10^{-1}$	$2.06 \cdot 10^{-5}$	$9.54 \cdot 10^{-9}$	$5.82 \cdot 10^{-1}$	$5.60 \cdot 10^{-2}$
$2.11 \cdot 10^{-2}$	$1.95 \cdot 10^{-5}$	1.67	$1.00 \cdot 10^{-1}$	$1.03 \cdot 10^{-5}$	$4.77 \cdot 10^{-9}$	$4.95 \cdot 10^{-1}$	$5.66 \cdot 10^{-2}$
$1.05 \cdot 10^{-2}$	$9.77 \cdot 10^{-6}$	1.45	$1.00 \cdot 10^{-1}$	$5.15 \cdot 10^{-6}$	$2.38 \cdot 10^{-9}$	$3.39 \cdot 10^{-1}$	$1.47 \cdot 10^{-2}$
$5.29 \cdot 10^{-3}$	$4.88 \cdot 10^{-6}$	1.29	$1.00 \cdot 10^{-1}$	$2.58 \cdot 10^{-6}$	$5.96 \cdot 10^{-10}$	$2.70 \cdot 10^{-1}$	$2.61 \cdot 10^{-2}$
$2.66 \cdot 10^{-3}$	$2.44 \cdot 10^{-6}$	1.18	$1.00 \cdot 10^{-1}$	$1.29 \cdot 10^{-6}$	$3.73 \cdot 10^{-11}$	$1.65 \cdot 10^{-1}$	$1.48 \cdot 10^{-2}$
$1.32 \cdot 10^{-3}$	$1.22 \cdot 10^{-6}$	$9.82 \cdot 10^{-1}$	$1.00 \cdot 10^{-1}$				

(a) $u^\dagger$

(b) u_α^δ for $\alpha = 1.25 \cdot 10^{-3}$

(c) u_α^δ for $\alpha = 10^{-4}$
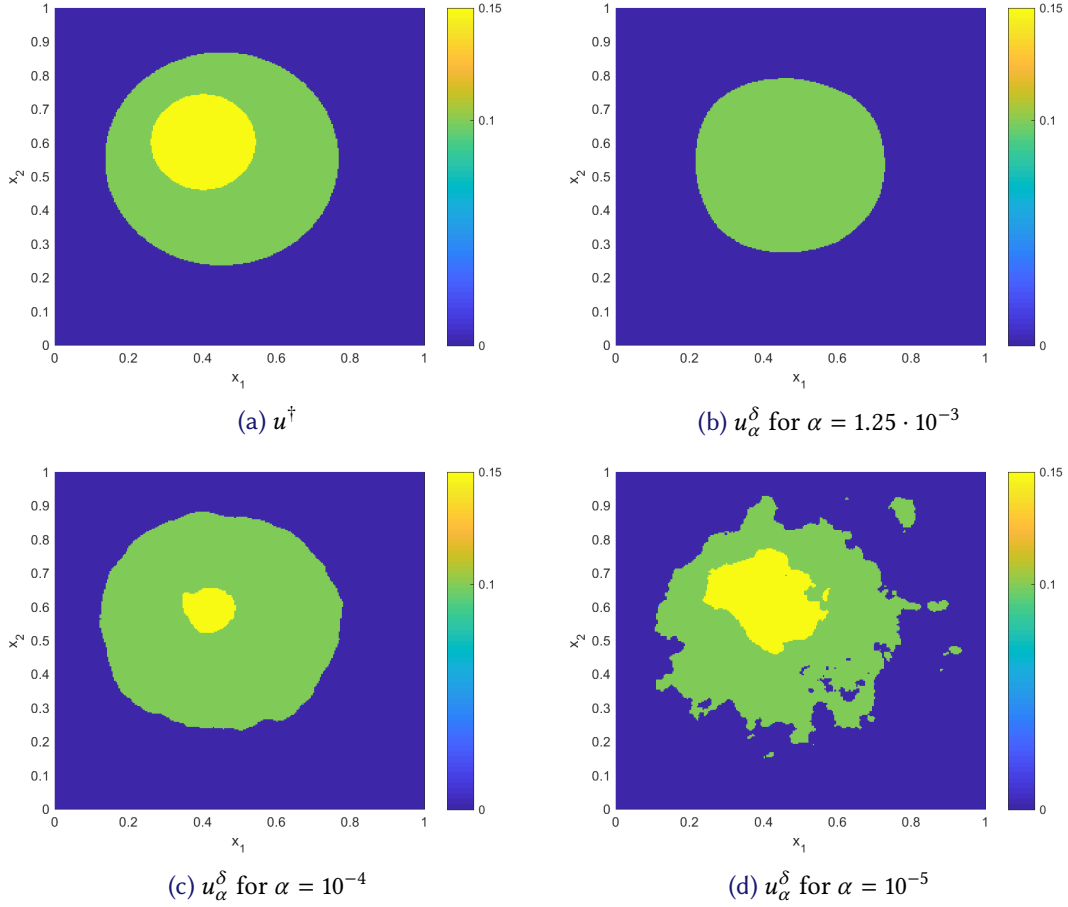
(d) u_α^δ for $\alpha = 10^{-5}$

Figure 5: True parameter $u^\dagger$ and reconstructions u_α^δ for $u_3 = 0.15$, $\delta \approx 7.59 \cdot 10^{-1}$, and different α

Figure 5b presents the reconstruction for the value $\alpha = 1.25 \cdot 10^{-3}$, where as before the volume corresponding to u_2 is reduced and the inner inclusion corresponding to u_3 is suppressed completely. If the parameter is chosen smaller as $\alpha = 10^{-4}$, however, the reconstruction of the outer volume is essentially correct, while the inner inclusion – although reduced – is also localized well; see Figure 5c. Visually, this value yields a better reconstruction than the one obtained by the discrepancy principle. The trade-off is a loss of spatial regularity, manifested in more irregular level lines, which becomes even more pronounced for smaller $\alpha = 10^{-5}$; see Figure 5d. This behavior is surprising insofar that the pointwise definition of the multi-bang penalty itself imposes no spatial regularity on the reconstruction at all; as is evident from (5.1), any regularity of the solution $\bar{u}$ is solely due to that of the level sets of $\bar{p}$ (which in this case has the regularity of a solution to a Poisson equation).

8 CONCLUSION

Reconstructions in inverse problems that take on values from a given discrete admissible set can be promoted via a convex penalty that leads to a convergent regularization method. While convergence rates can be shown with respect to the usual Bregman distance, if the true parameter to be reconstructed takes on values only from the admissible set, the convergence (albeit without rates) is actually pointwise. A semismooth Newton method allows the efficient and robust computation of Tikhonov minimizers.

This work can be extended in several directions. First, Figure 5 demonstrates that regularization parameters chosen according to the discrepancy principle are not optimal with respect to the visual reconstruction quality. This motivates the development of new, heuristic, parameter choice rules that are adapted to the discrete-valued, pointwise, nature of the multi-bang penalty. It would also be interesting to investigate whether an active set condition in the spirit of [28, 29] based on (4.1) can be used to obtain strong or pointwise convergence rates. A natural further step is the extension to nonlinear parameter identification problems, making use of the results of [10]. Finally, Figures 5c and 5d suggest combining the multi-bang penalty with a total variation penalty to also promote regularity of the level lines of the reconstruction. The resulting problem is challenging both analytically and numerically, but would open up the possibility of application to electrical impedance tomography, which can be formulated as parameter identification problem for the diffusion coefficient in an elliptic equation.

REFERENCES

[1] Bae & Tai, Graph cut optimization for the piecewise constant level set method applied to multiphase image segmentation, in: *Scale Space and Variational Methods in Computer Vision: Second International Conference, SSVM 2009, Voss, Norway, June 1-5, 2009. Proceedings*, Springer Berlin Heidelberg, 2009, 1–13, DOI: 10.1007/978-3-642-02256-2_1.

[2] Barbu & Precupanu, *Convexity and Optimization in Banach Spaces*, Springer, Dordrecht, 2012, DOI: 10.1007/978-94-007-2247-7.

[3] Bauschke & Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2011, DOI: 10.1007/978-1-4419-9467-7.

[4] Bergounioux & Tröltzsch, Optimality conditions and generalized bang-bang principle for a state-constrained semilinear parabolic problem, *Numer. Funct. Anal. Optim.* 17 (1996), 517–536, DOI: 10.1080/01630569608816708.

[5] Bregman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, *USSR Computational Mathematics and Mathematical Physics* 7 (1967), 200–217.

[6] Burger & Osher, Convergence rates of convex variational regularization, *Inverse Problems* 20 (2004), 1411, DOI: 10.1088/0266-5611/20/5/005.

[7] Cai, Chan & Zeng, A Two-Stage Image Segmentation Method Using a Convex Variant of the Mumford–Shah Model and Thresholding, *SIAM Journal on Imaging Sciences* 6 (2013), 368–390, DOI: 10.1137/120867068.

[8] Clason, Ito & Kunisch, A convex analysis approach to optimal controls with switching structure for partial differential equations, *ESAIM: Control, Optimisation and Calculus of Variations* 22 (2016), 581–609, DOI: 10.1051/cocv/2015017.

[9] Clason & Kunisch, Multi-bang control of elliptic systems, *Annales de l'Institut Henri Poincaré (C) Analyse Non Linéaire* 31 (2014), 1109–1130, DOI: 10.1016/j.anihpc.2013.08.005.

[10] Clason & Kunisch, A convex analysis approach to multi-material topology optimization, *ESAIM: Mathematical Modelling and Numerical Analysis* 50 (2016), 1917–1936, DOI: 10.1051/m2an/2016012.

[11] Clason, Tameling & Wirth, Vector-valued multibang control of differential equations, *arXiv* 1611 (2016), URL: http://www.arxiv.org/abs/1611.07853.

[12] Ekeland & Témam, *Convex Analysis and Variational Problems*, SIAM, 1999, DOI: 10.1137/1.9781611971088.

[13] Flemming & Hofmann, Convergence rates in constrained Tikhonov regularization: Equivalence of projected source conditions and variational inequalities, *Inverse Problems* 27 (2011), 085001, DOI: 10.1088/0266-5611/27/8/085001.

[14] Goldluecke & Cremers, Convex relaxation for multilabel problems with product label spaces, in: *Computer Vision – ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part V*, Springer Berlin Heidelberg, 2010, 225–238, DOI: 10.1007/978-3-642-15555-0_17.

[15] Hofmann, Kaltenbacher, Pöschl & Scherzer, A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators, *Inverse Problems* 23 (2007), 987–1010, DOI: 10.1088/0266-5611/23/3/009.

[16] Ishikawa, Exact optimization for Markov Random Fields with convex priors, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25 (2003), 1333–1336, DOI: 10.1109/TPAMI.2003.1233908.

[17] Ito & Jin, *Inverse Problems: Tikhonov Theory and Algorithms*, World Scientific, 2014, DOI: 10.1142/9789814596206_0001.

[18] Ito & Kunisch, *Lagrange Multiplier Approach to Variational Problems and Applications*, SIAM, 2008, DOI: 10.1137/1.9780898718614.

[19] Lellmann & Schnörr, Continuous Multiclass Labeling Approaches and Algorithms, *SIAM Journal on Imaging Sciences* 4 (2011), 1049–1096, DOI: 10.1137/100805844.

[20] Resmerita, Regularization of ill-posed problems in Banach spaces: convergence rates, *Inverse Problems* 21 (2005), 1303, DOI: 10.1088/0266-5611/21/4/007.

[21] Scherzer, Grasmair, Grossauer, Haltmeier & Lenzen, *Variational Methods in Imaging*, Springer, 2009, DOI: 10.1007/978-0-387-69277-7.

[22] Schirotzek, *Nonsmooth Analysis*, Springer, 2007, DOI: 10.1007/978-3-540-71333-3.

[23] Schuster, Kaltenbacher, Hofmann & Kazimierski, *Regularization methods in Banach spaces*, De Gruyter, 2012, DOI: 10.1515/9783110255720.

[24] Tröltzsch, A minimum principle and a generalized bang-bang principle for a distributed optimal control problem with constraints on control and state, *Z. Angew. Math. Mech.* 59 (1979), 737–739, DOI: 10.1002/zamm.19790591208.

[25] Tröltzsch, *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*, American Mathematical Society, 2010, DOI: 10.1090/gsm/112.

[26] Ulbrich, *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*, SIAM, 2011, DOI: 10.1137/1.9781611970692.

[27] Vese & Chan, A Multiphase Level Set Framework for Image Segmentation Using the Mumford and Shah Model, *International Journal of Computer Vision* 50 (2002), 271–293, DOI: 10.1023/A:1020874308076.

[28] Wachsmuth & Wachsmuth, Regularization error estimates and discrepancy principle for optimal control problems with inequality constraints, *Control Cybernet.* 40 (2011), 1125–1158.

[29] Wachsmuth & Wachsmuth, Convergence and regularization results for optimal control problems with sparsity functional, *ESAIM Control Optim. Calc. Var.* 17 (2011), 858–886, DOI: 10.1051/cocv/2010027.