ruhr.paD
UA Ruhr Zentrum für
partielle Differentialgleichungen

Error estimates for the approximation of a discrete-valued optimal control problem

Ch. Clason, T.B.T. Do and F. Pörner

Preprint 2018-04

ERROR ESTIMATES FOR THE APPROXIMATION OF A DISCRETE-VALUED OPTIMAL CONTROL PROBLEM

Christian Clason[*] Thi Bich Tram Do[*] Frank Pörner[†]

March 12, 2018

Abstract This work is concerned with optimal control problems with a multibang promoting objective functional, where the objective functional consists of a tracking-type functional and an additional regularization functional that promotes optimal control taking values from a given discrete set pointwise almost everywhere. Under a regularity condition on the set where these discrete values are attained, error estimates for the Moreau–Yosida approximation (which allows its solution by a semismooth Newton method) and the discretization of the problem are derived. Numerical results support the theoretical findings.

1 INTRODUCTION

We consider linear-quadratic optimal control problems where the optimal control is only allowed to take values at discrete values $u_1 < \cdots < u_d \in \mathbb{R}$ with $d \in \mathbb{N}$. Such problems occur, e.g., in topology optimization, nondestructive testing or medical imaging; a similar task also arises as a sub-step in segmentation or labeling problems in image processing. However, such problems are inherently nonconvex and, more importantly, not weakly lower semi-continuous and hence cannot be treated by standard techniques. A classical remedy is convex relaxation, where the nonconvex constraint $u(x) \in \{u_1, \ldots, u_d\}$ is replaced by the convex constraint $u(x) \in [u_1, u_d]$, but this leads to ignoring the intermediate parameter values. In [3, 5–8], it was therefore proposed to promote all desired control values using a convex *multibang* penalty

$$G(u) : L^2(\Omega) \to \mathbb{R}, \qquad u \mapsto \int_\Omega g(u(x))\,\mathrm{d}x,$$

for a suitable convex integrand $g : \mathbb{R} \to \mathbb{R}$ with a polyhedral epigraph whose vertices correspond to the desired control values $u_1, \ldots, u_d$. We thus consider the *multibang control problem*

$$\min_{u \in L^2(\Omega)} \frac{1}{2}\|Ku - z\|_Y^2 + \alpha G(u) \tag{1.1}$$

[*] Faculty of Mathematics, University Duisburg-Essen, 45117 Essen, Germany (christian.clason@uni-due.de, tram.do@uni-due.de)

[†] Department of Mathematics, University of Würzburg, Emil-Fischer-Str. 40, 97074 Würzburg, Germany (frank.poerner@mathematik.uni-wuerzburg.de)

with $\alpha > 0, z \in Y$ for a Hilbert space Y, and $K : L^2(\Omega) \longrightarrow Y$ a linear and continuous operator (e.g., the solution operator for a linear elliptic partial differential equation). Just as in L^1 regularization for sparsity (and in linear optimization), it can be expected that minimizers are found at the vertices of G, thus yielding the desired structure. Furthermore, it was shown in [3, 4, 7] that this leads to a primal-dual optimality system that can be solved by a superlinearly convergent semismooth Newton method in function space [14, 22] if a suitable Moreau–Yosida approximation (of the Fenchel conjugate G^*, see Proposition 2.2 below) is introduced. It turns out that this approximation can be expressed in primal form as

$$\min_{u \in L^2(\Omega)} \frac{1}{2}\|Ku - z\|_Y^2 + \alpha G(u) + \gamma \|u\|_{L^2(\Omega)}^2 \tag{1.2}$$

for a parameter $\gamma > 0$. We remark that this approach (i.e., applying the approximation to G^* instead of G) does not destroy the non-differentiability of G and hence preserves the structural properties of (1.1). Standard lower semicontinuity techniques can then be applied to show that the solutions to (1.2) converge weakly to the solution to (1.1) as $\gamma \longrightarrow 0$; see [7, § 4.1]. The aim of this paper is to establish strong convergence and in particular approximation error estimates for $\|\bar{u} - u_\gamma\|_{L^2(\Omega)}$.

Let us recall some literature and already known results. For the case $d = 2$ we obtain the minimization problem

$$\min_{u_1 < u < u_2} \frac{1}{2}\|Ku - z\|_Y^2. \tag{1.3}$$

and if the associated adjoint state $\bar{p}(x) \neq 0$ almost everywhere, it is well-known that $\bar{u}$ exhibits a bang-bang structure, i.e. $\bar{u}(x) \in \{u_1, u_2\}$ almost everywhere. This problem has been studied intensively in the literature, see [20, 21, 23, 25, 26] and the references therein. Note that this list is far away from being complete. For this problem a structural assumption has been established in [25, 26], which controls the behavior of the adjoint state around a singular set and guarantees that the optimal control $\bar{u}$ exhibits a bang-bang structure. Using this assumption, error estimates for the approximation of (1.3) can be proven; see [25]. A related question is the Moreau–Yosida approximation of state constraints; see [10, 11].

If $d = 3$ and $u_1 < u_2 = 0 < u_2$, the problem (1.1) resembles the minimization problem

$$\min_{u_1 < u < u_2} \frac{1}{2}\|Ku - z\|_Y^2 + \alpha\|u\|_{L^1(\Omega)}, \tag{1.4}$$

see, e.g., [20]. The structural assumption used to prove error rates for the approximation of (1.3) can be generalized to problem (1.4). Again, approximation error estimates can be proven; see [23, 25, 26] and the reference therein.

We will generalize this structural assumption to the multibang control problem (1.1). We will show that this assumption is sufficient to guarantee that an optimal control $\bar{u}$ of (1.1) satisfies $\bar{u}(x) \in \{u_1, \dots, u_d\}$ for almost all $x \in \Omega$. Furthermore, we will use this condition to prove approximation error estimates of the form

$$\|\bar{u} - u_\gamma\|_{L^2(\Omega)} = O\left(\gamma^{-\frac{\kappa}{2}}\right)$$

with a constant $\kappa > 0$ depending only on the structural assumption.

The paper is organized as follows. In Section 2 we recall some preliminary results which are needed for the convergence analysis. Our structural assumption is introduced in Section 3 and used to derive the approximation error estimates. This is also the main result of this paper. In Section 4, we establish discretization error estimates under our structural assumption. We introduce an active set method for the solution of (1.2) and show its equivalence to a semismooth Newton method in Section 5. Finally, numerical results to support our theoretical findings can be found in Section 6.

2 PRELIMINARY RESULTS

Let $u_1 < u_2 < \cdots < u_d$ be some given real numbers with $d \geq 2$, and let $\Omega \subset \mathbb{R}^n$ be a bounded domain. Following [3, 5–7], we define the piecewise linear function

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

As the pointwise supremum of affine functions, g is convex and continuous on the interior of its domain $\operatorname{dom}(g) = [u_1, u_d]$. Hence, the corresponding integral functional

$$G : L^2(\Omega) \to \mathbb{R}, \qquad u \mapsto \int_\Omega g(u(x))\,\mathrm{d}x,$$

is proper, convex and weakly lower semicontinuous as well; see, e.g., [2, Proposition 2.53].

We now consider the problem

$$\min_{u \in L^2(\Omega)} \frac{1}{2}\|Ku - z\|_Y^2 + \alpha G(u) \tag{2.1}$$

with a parameter $\alpha > 0$. Standard semi-continuity methods then yield existence of a minimizer $\bar{u}$, which is unique if K is injective; cf. [7]. We will later impose an condition which guarantees that $\bar{u}$ exhibits a multibang structure, i.e., $\bar{u}(x) \in \{u_1, \dots, u_d\}$ for almost every $x \in \Omega$.

Let us further define the set

$$U_{\text{ad}} := \{u \in L^2(\Omega) : u_1 \leq u(x) \leq u_d\} = \operatorname{co}\left\{u \in L^2(\Omega) : u(x) \in \{u_1, \dots, u_d\}\right\}.$$

It is clear that (2.1) is equivalent to the problem

$$\min_{u \in U_{\text{ad}}} \frac{1}{2}\|Ku - z\|_Y^2 + \alpha G(u). \tag{P}$$

We will use this equivalent formulation to derive variational inequalities which will be useful in the convergence analysis. Standard convex analysis techniques then yield primal–dual optimality conditions; see, e.g.,[3, 7].

Proposition 2.1. *Define the sets*

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

Let $\bar u \in U_{\text{ad}}$ with associated adjoint state $\bar p := K^*(z - K\bar u)$. Then $\bar u$ is a solution to (P) *if and only if*

$$\bar u(x) \in \begin{cases} \{u_i\} & \textit{if}\ \bar p(x) \in Q_i \quad 1 \le i \le d, \\ [u_i, u_{i+1}] & \textit{if}\ \bar p(x) \in Q_{i,i+1} \quad 1 \le i < d. \end{cases} \tag{2.2}$$

It is clear that the optimal solution $\bar u$ is uniquely determined by the adjoint state on the sets Q_i. We see furthermore that $\bar u(x) \in \{u_1, \dots, u_d\}$ almost everywhere on Ω if $\operatorname{meas}(Q_{i,i+1}) = 0$ for all $1 \le i < d$. Hence $\bar u$ has a multibang structure in this case. In the following, we will make use of this relation to construct a suitable regularity condition on these sets.

We next introduce the Moreau–Yosida approximation of (P) with a regularization parameter $\gamma > 0$,

$$\min_{u \in U_{\text{ad}}} \frac{1}{2}\|Ku - z\|_Y^2 + \alpha G(u) + \frac{\gamma}{2}\|u\|_{L^2(\Omega)}^2. \tag{P_γ}$$

As for (P), arguments from convex analysis lead to the following optimality conditions; see [3, 7].

Proposition 2.2. *Define the sets*

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

Let $u_\gamma \in U_{\text{ad}}$ with associated adjoint state $p_\gamma := K^*(z - Ku_\gamma)$. Then u_γ is a solution to (P_γ) *if and only if*

$$u_\gamma(x) \in \begin{cases} \{u_i\} & \textit{if}\ p_\gamma(x) \in Q_i^\gamma \quad 1 \le i \le d, \\ \frac{1}{\gamma}\left(p_\gamma(x) - \frac{\alpha}{2}(u_i + u_{i+1}) \right) & \textit{if}\ p_\gamma(x) \in Q_{i,i+1} \quad 1 \le i < d. \end{cases} \tag{2.3}$$

We remark that (2.3) are the explicit pointwise characterization of $u_\gamma \in (\partial G^*)_\gamma(p_\gamma)$, where $(\partial G^*)_\gamma$ denotes the Yosida approximation of the convex subdifferential (which coincides with the Fréchet derivative of the Moreau envelope) of the Fenchel conjugate of G, which justifies the term *Moreau–Yosida approximation*; see, e.g., [3, § 4.1].

We can also derive purely primal first-order optimality conditions for (P) and (P_γ) in terms of variational inequalities using standard arguments as in, e.g., [21, Thm. 2.22].

Proposition 2.3. *Let $\bar{u}$ and u_γ be solutions of (P) and (P_γ) with associated adjoint states $\bar{p} := K^*(z - K\bar{u})$ and $p_\gamma := K^*(z - Ku_\gamma)$, respectively. Then,*

$$\begin{aligned}(-\bar{p}, u - \bar{u})_{L^2(\Omega)} + \alpha G'(\bar{u}; u - \bar{u}) \geq 0 \quad &\textit{for all } u \in U_{\text{ad}},\\ \left(-p_\gamma + \gamma u_\gamma, u - u_\gamma\right)_{L^2(\Omega)} + \alpha G'(u_\gamma; u - u_\gamma) \geq 0 \quad &\textit{for all } u \in U_{\text{ad}}.\end{aligned}$$

Here, $G'(\bar{u}; u - \bar{u})$ denotes the directional derivative of G at $\bar{u}$ in direction $u - \bar{u}$, which will be characterized in the following lemma. Note that for $\bar{u}, u \in U_{\text{ad}}$ we have $u - \bar{u} \in T_{U_{\text{ad}}}(\bar{u})$ for

$$T_{U_{\text{ad}}}(u) := \left\{ v \in L^2(\Omega) : v(x) \begin{cases} \geq 0 & \text{if } u(x) = u_1 \\ \leq 0 & \text{if } u(x) = u_d \end{cases} \right\},$$

i.e., the tangential cone to U_{ad} in the point u. It thus suffices to consider directional derivatives for directions in $T_{U_{\text{ad}}}$, which helps to avoid unnecessary case distinctions in the proof.

Lemma 2.4. *Let $u \in U_{\text{ad}}$ and define the sets*

$$\begin{aligned} S_i &:= \{x \in \Omega : u(x) = u_i\}, \quad i = 1, \ldots, d,\\ T_i &:= \{x \in \Omega : u_i < u(x) < u_{i+1}\}, \quad i = 1, \ldots, d-1.\end{aligned}$$

The directional derivative of G in direction $v \in T_{U_{\text{ad}}}(u)$ is then given as

$$\begin{aligned} G'(u; v) = &\sum_{i=1}^{d-1} \int_{T_i} \frac{1}{2}(u_i + u_{i+1})v(x)\, \mathrm{d}x \\ &+ \sum_{i=1}^{d} \left[\int_{S_i \cap \{v \geq 0\}} \frac{1}{2}(u_i + u_{i+1})v(x)\, \mathrm{d}x + \int_{S_i \cap \{v < 0\}} \frac{1}{2}(u_{i-1} + u_i)v(x)\, \mathrm{d}x \right].\end{aligned}$$

Proof. We use the definition of the directional derivative and of the sets S_i and T_i to obtain

$$\begin{aligned} G'(u; v) &:= \lim_{\rho \to 0} \frac{1}{\rho} \left(G(u + \rho v) - G(u)\right) \\ &= \lim_{\rho \to 0} \frac{1}{\rho} \left[\sum_{i=1}^{d-1} \int_{T_i} (g(u(x) + \rho v(x)) - g(u(x)))\, \mathrm{d}x + \sum_{i=1}^{d} \int_{S_i} (g(u(x) + \rho v(x)) - g(u(x)))\, \mathrm{d}x \right].\end{aligned}$$

We now have to differentiate between several cases.

(i) First, assume that $x \in T_i$ with $1 \leq i \leq d-1$. For ρ small enough we then get $u(x) + \rho v(x) \in T_i$. Hence we obtain

$$\begin{aligned} g(u(x) + \rho v(x)) - g(u(x)) &= \frac{1}{2}((u_i + u_{i+1})(u(x) + \rho v(x)) - u_i u_{i+1}) \\ &\quad - \frac{1}{2}((u_i + u_{i+1})u(x) - u_i u_{i+1}) \\ &= \frac{\rho}{2}(u_i + u_{i+1})v(x),\end{aligned}$$

which yields

$$\lim_{\rho\to 0}\int\limits_{T_i}(g(u(x)+\rho v(x))-g(u(x)))\,\mathrm{d}x=\int\limits_{T_i}\frac{1}{2}(u_i+u_{i+1})v(x)\,\mathrm{d}x.$$

(ii) Now assume that $x\in S_i$ with $1<i<d$. Proposition 2.1 then implies that $u(x)=u_i$. Here we have to further differentiate between three cases.

$v(x)=0$: Here we obtain $u(x)+\rho v(x)=u(x)$, leading to

$$g(u(x)+\rho v(x))-g(u(x))=0.$$

$v(x)>0$: Here we obtain $u(x)+\rho v(x)\in T_i$ for ρ small enough, leading to

$$g(u(x)+\rho v(x))-g(u(x))=\frac{\rho}{2}(u_i+u_{i+1})v(x).$$

$v(x)<0$: Here we obtain $u(x)+\rho v(x)\in T_{i-1}$, leading to

$$\begin{aligned}g(u(x)+\rho v(x))-g(u(x))&=\frac{1}{2}((u_{i-1}+u_i)(u_i+\rho v(x))-u_{i-1}u_i)\\&\quad-\frac{1}{2}((u_i+u_{i+1})u_i-u_iu_{i+1})\\&=\frac{\rho}{2}(u_{i-1}+u_i)v(x).\end{aligned}$$

Combining all three cases yields

$$\begin{aligned}&\lim_{\rho\to 0}\frac{1}{\rho}\int\limits_{S_i}(g(u(x)+\rho v(x))-g(u(x)))\,\mathrm{d}x\\&\qquad=\int\limits_{S_i\cap\{v\geq 0\}}\frac{1}{2}(u_i+u_{i+1})v(x)\,\mathrm{d}x+\int\limits_{S_i\cap\{v<0\}}\frac{1}{2}(u_{i-1}+u_i)v(x)\,\mathrm{d}x.\end{aligned}$$

(iii) We are left with the special cases $x\in S_i$ for $i=1$ and $i=d$. We only consider the case $i=1$ as the case $i=d$ is similar. Hence we assume $x\in S_1$, which implies $u(x)=u_1$. Since $v\in T_{U_{\mathrm{ad}}}(u)$, we have that $v(x)\geq 0$. If $v(x)>0$, we obtain for ρ small enough that $u(x)+\rho v(x)\in T_1$ holds, leading to

$$g(u(x)+\rho v(x))-g(u(x))=\frac{\rho}{2}(u_1+u_2)v(x)$$

and similar if $v(x)=0$. This leads to

$$\lim_{\rho\to 0}\frac{1}{\rho}\int\limits_{S_1}(g(u(x)+\rho v(x))-g(u(x)))\,\mathrm{d}x=\int\limits_{S_1}\frac{1}{2}(u_1+u_2)v(x)\,\mathrm{d}x.$$

A similar argument for the remaining case $i=d$ finishes the proof. □

3 REGULARITY ASSUMPTION AND ERROR ESTIMATES

We now extend the active set condition from [25, 26] to the multibang control problem. From Proposition 2.1, we see that the optimal control $\bar{u}$ is not uniquely determined by the adjoint state $\bar{p}$ on the *singular sets* $Q_{i,i+1}$. We therefore need to control the way in which $\bar{p}$ "detaches" from these sets. This motivates the following assumption.

Assumption REG. For the solution $\bar{u}$ to (P) with adjoint state $\bar{p} = K^*(z - K\bar{u})$ there exists a constant $c > 0$ and $\kappa > 0$ such that

$$\operatorname{meas}\left(\bigcup_{i=1}^{d-1}\left\{x \in \Omega : \left|\bar{p}(x) - \frac{\alpha}{2}(u_i + u_{i+1})\right| < \varepsilon\right\}\right) \leq c\varepsilon^{\kappa}$$

holds for all $\varepsilon > 0$ small enough.

Note that if $\bar{u}$ satisfies this assumption, the sets $Q_{i,i+1}$ have Lebesgue measure zero. Hence, $\bar{u}$ is multibang by Proposition 2.1. In addition, we have the following result, which is a direct consequence of $\operatorname{meas}(Q_{i,i+1}) = 0$.

Lemma 3.1. *Assume $\bar{u}$ satisfies Assumption REG. Then $\bar{p}(x) \in Q_i$ if and only if $\bar{u}(x) = u_i$ holds almost everywhere in Ω.*

Following [9, Lemma 1.3], we can derive a sufficient condition for Assumption REG.

Theorem 3.2. *Suppose that the adjoint state $\bar{p} \in C^1(\bar{\Omega})$ and satisfies*

$$\min_{x \in K_i} |\nabla p(x)| > 0 \qquad \text{for all } i = 1, \ldots, d-1, \text{ for}$$

where

$$K_i := \left\{x \in \bar{\Omega} : p(x) = \frac{\alpha}{2}(u_i + u_{i+1})\right\}.$$

Then Assumption REG holds with $\kappa = 1$.

Proof. Define for $t \in \mathbb{R}$ the level sets $F_t := \{x \in \bar{\Omega} : p(x) = t\}$. Now we use a continuity argument to obtain constants $\varepsilon_0, c_0, C > 0$ such that for all $|t - \frac{\alpha}{2}(u_i + u_{i+1})| \leq \varepsilon_0$ and all $1 \leq i < d$ there holds

$$|\nabla p(x)| \geq c_0, \quad \mathcal{H}^{n-1}(F_t) \leq C,$$

where $\mathcal{H}^{n-1}$ is the $(n-1)$-dimensional Hausdorff measure. We now use the co-area formula

$$\int_{\Omega} h(x)|\nabla p(x)|\,\mathrm{d}x = \int_{-\infty}^{\infty}\left(\int_{p^{-1}(t)} h(x)\mathrm{d}\mathcal{H}^{n-1}(x)\right)\mathrm{d}t$$

with the function

$$h(x) := \mathbb{1}_{E_i}, \quad E_i := \left\{x \in \Omega : \left|p(x) - \frac{\alpha}{2}(u_i + u_{i+1})\right| \leq \varepsilon\right\},$$

to obtain for all $1 \leq i < d$ and $0 < \varepsilon \leq \varepsilon_0$ that

$$c_0 \operatorname{meas}(E_i) \leq \int\limits_{E_i} |\nabla p(x)| \,\mathrm{d}x = \int\limits_{-\varepsilon}^{\varepsilon} \mathcal{H}^{n-1}\left(F_{t-\frac{\alpha}{2}(u_i+u_{i+1})}\right) \mathrm{d}t \leq 2C\varepsilon$$

holds. Since this holds for all $1 \leq i < d$, the Assumption REG now follows with $\kappa = 1$. □

We now establish error estimates for the approximation (P_γ) of (P). For this purpose, we first derive a stronger version of Proposition 2.3. The next result, which is similar to ones in [18, 19], is the most important tool in the convergence analysis.

Lemma 3.3. *Assume that the solution $\bar{u}$ to (P) satisfies Assumption REG. Then,*

$$(-\bar{p}, u - \bar{u})_{L^2(\Omega)} + \alpha G'(\bar{u}; u - \bar{u}) \geq c_A \|u - \bar{u}\|_{L^1(\Omega)}^{1+\frac{1}{\kappa}} \qquad \forall u \in U_{\mathrm{ad}}$$

with a constant $c_A > 0$.

Proof. First, recall that Assumption REG implies that $\bar{u}$ has a multibang structure. Furthermore, using Lemma 3.1 we obtain with the definition of Q_i and S_i in Proposition 2.1 and Lemma 2.4, respectively, that $\bar{u}(x) \in S_i$ if and only if $\bar{p}(x) \in Q_i$. Now we use Lemma 2.4 and the fact that $u - \bar{u} \in T_{U_{\mathrm{ad}}}(\bar{u})$ to compute

$$\begin{aligned}
&(-\bar{p}, u - \bar{u})_{L^2(\Omega)} + \alpha G'(\bar{u}; u - \bar{u}) \\
&\qquad = \int\limits_{\{\bar{p} \in Q_1\}} \left(-\bar{p}(x) + \frac{1}{2}(u_1 + u_2)\right)(u(x) - \bar{u}(x)) \,\mathrm{d}x \\
&\qquad\quad + \int\limits_{\{\bar{p} \in Q_d\}} \left(-\bar{p}(x) + \frac{1}{2}(u_{d-1} + u_d)\right)(u(x) - \bar{u})(x) \,\mathrm{d}x \\
&\qquad\quad + \sum_{i=2}^{d-1} \int\limits_{\{\bar{p} \in Q_i\} \cap \{u - \bar{u} \geq 0\}} \left(-\bar{p}(x) + \frac{\alpha}{2}(u_i + u_{i+1})\right)(u(x) - \bar{u}(x)) \,\mathrm{d}x \\
&\qquad\quad + \sum_{i=2}^{d-1} \int\limits_{\{\bar{p} \in Q_i\} \cap \{u - \bar{u} < 0\}} \left(-\bar{p}(x) + \frac{\alpha}{2}(u_{i-1} + u_i)\right)(u(x) - \bar{u}(x)) \,\mathrm{d}x.
\end{aligned}$$

Here we have abbreviated the sets $\{\bar{p} \in Q_1\} := \{x \in \Omega : \bar{p}(x) \in Q_1\}$ and similar for the other sets. Recall that by definition, $\bar{p}(x) \in Q_1$ implies that $-\bar{p}(x) + \frac{\alpha}{2}(u_1 + u_2) > 0$. Furthermore, we know that $\bar{u}(x) = u_1$, leading to $u(x) - \bar{u}(x) = u(x) - u_1 \geq 0$. We similarly obtain on Q_d that $-\bar{p}(x) + \frac{\alpha}{2}(u_{d-1} + u_d) < 0$ and $u(x) - \bar{u}(x) = u(x) - u_d \leq 0$. Finally, if $\bar{p}(x) \in Q_i$ for $1 < i < d$, we obtain that

$$\frac{\alpha}{2}(u_{i-1} + u_i) < \bar{p}(x) < \frac{\alpha}{2}(u_i + u_{i+1}),$$

which leads to

$$-\bar{p}(x) + \frac{\alpha}{2}(u_i + u_{i+1}) > 0 \quad \text{and} \quad -\bar{p}(x) + \frac{\alpha}{2}(u_{i-1} + u_i) < 0.$$

This allows us to write

$$\begin{aligned}
(-\bar{p}, u-\bar{u})_{L^2(\Omega)} &+ \alpha G'(\bar{u}; u-\bar{u}) \\
&= \int\limits_{\{\bar{p}\in Q_1\}} \left|-\bar{p}(x) + \frac{1}{2}(u_1+u_2)\right| |u(x)-\bar{u}(x)| \,\mathrm{d}x \\
&+ \int\limits_{\{\bar{p}\in Q_d\}} \left|-\bar{p}(x) + \frac{1}{2}(u_{d-1}+u_d)\right| |u(x)-\bar{u}(x)| \,\mathrm{d}x \\
&+ \sum_{i=2}^{d-1} \int\limits_{\{\bar{p}\in Q_i\}\cap\{u-\bar{u}\geq 0\}} \left|-\bar{p}(x) + \frac{\alpha}{2}(u_i+u_{i+1})\right| |u(x)-\bar{u}(x)| \,\mathrm{d}x \\
&+ \sum_{i=2}^{d-1} \int\limits_{\{\bar{p}\in Q_i\}\cap\{u-\bar{u}< 0\}} \left|-\bar{p}(x) + \frac{\alpha}{2}(u_{i-1}+u_i)\right| |u(x)-\bar{u}(x)| \,\mathrm{d}x.
\end{aligned}$$

Now let $\varepsilon > 0$ and consider the set

$$Q_1^\varepsilon := \left\{q : q \leq \frac{\alpha}{2}(u_1+u_2) - \varepsilon\right\} \subset Q_1.$$

Let $\bar{p}(x) \in Q_1^\varepsilon$. Together with $-\bar{p}(x) + \frac{\alpha}{2}(u_1+u_2) > 0$, this implies that

$$\left|-\bar{p}(x) + \frac{\alpha}{2}(u_1+u_2)\right| = -\bar{p}(x) + \frac{\alpha}{2}(u_1+u_2) \geq \varepsilon,$$

leading to

$$\begin{aligned}
\int\limits_{\{\bar{p}\in Q_1\}} \left|-\bar{p} + \frac{1}{2}(u_1+u_2)\right| |u-\bar{u}| \,\mathrm{d}x &\geq \int\limits_{\{\bar{p}\in Q_1^\varepsilon\}} \left|-\bar{p} + \frac{1}{2}(u_1+u_2)\right| |u-\bar{u}| \,\mathrm{d}x \\
&\geq \varepsilon \int\limits_{\{\bar{p}\in Q_1^\varepsilon\}} |u-\bar{u}| \,\mathrm{d}x.
\end{aligned}$$

We similarly define

$$Q_d^\varepsilon := \left\{q \geq \frac{\alpha}{2}(u_{d-1}+u_d) + \varepsilon\right\},$$

leading to

$$\int\limits_{\{\bar{p}\in Q_d\}} \left|-\bar{p}(x) + \frac{1}{2}(u_{d-1}+u_d)\right| |u(x)-\bar{u}(x)| \,\mathrm{d}x \geq \varepsilon \int\limits_{\{\bar{p}\in Q_d^\varepsilon\}} |u(x)-\bar{u}(x)| \,\mathrm{d}x,$$

as well as for $1 < i < d$

$$Q_i^\varepsilon := \left\{\varepsilon + \frac{\alpha}{2}(u_{i-1}+u_i) \leq q \leq \frac{\alpha}{2}(u_i+u_{i+1}) - \varepsilon\right\} \subset Q_i.$$

The latter leads to

$$\begin{aligned}\left|-\bar{p}(x) + \frac{\alpha}{2}(u_i + u_{i+1})\right| &= -\bar{p}(x) + \frac{\alpha}{2}(u_i + u_{i+1}) \geq \varepsilon,\\ \left|-\bar{p}(x) + \frac{\alpha}{2}(u_{i-1} + u_i)\right| &= \ \ \bar{p}(x) - \frac{\alpha}{2}(u_{i-1} + u_i) \geq \varepsilon\end{aligned}$$

and therefore

$$\begin{aligned}&\int_{\{\bar{p}\in Q_i\}\cap\{u-\bar{u}\geq 0\}} \left|-\bar{p}(x) + \frac{\alpha}{2}(u_i + u_{i+1})\right| |u(x) - \bar{u}(x)|\,\mathrm{d}x\\ &\qquad + \int_{\{\bar{p}\in Q_i\}\cap\{u-\bar{u}<0\}} \left|-\bar{p}(x) + \frac{\alpha}{2}(u_{i-1} + u_i)\right| |u(x) - \bar{u}(x)|\,\mathrm{d}x\\ &\qquad \geq \varepsilon \int_{\{\bar{p}\in Q_i^\varepsilon\}\cap\{u-\bar{u}\geq 0\}} |u(x) - \bar{u}(x)|\,\mathrm{d}x + \varepsilon \int_{\{\bar{p}\in Q_i^\varepsilon\}\cap\{u-\bar{u}<0\}} |u(x) - \bar{u}(x)|\,\mathrm{d}x\\ &\qquad = \varepsilon \int_{\{\bar{p}\in Q_i^\varepsilon\}} |u(x) - \bar{u}(x)|\,\mathrm{d}x.\end{aligned}$$

We now combine all these estimates to obtain

$$\begin{aligned}&(-\bar{p}, u - \bar{u})_{L^2(\Omega)} + \alpha G'(\bar{u}; u - \bar{u})\\ &\qquad \geq \varepsilon \sum_{i=1}^{d} \int_{\{\bar{p}\in Q_i^\varepsilon\}} |u(x) - \bar{u}(x)|\,\mathrm{d}x\\ &\qquad = \varepsilon \sum_{i=1}^{d} \left(\int_{\{\bar{p}\in Q_i\}} |u(x) - \bar{u}(x)|\,\mathrm{d}x - \int_{\{\bar{p}\in Q_i\}\setminus\{\bar{p}\in Q_i^\varepsilon\}} |u(x) - \bar{u}(x)|\,\mathrm{d}x \right)\\ &\qquad = \varepsilon \|u - \bar{u}\|_{L^1(\Omega)} - \varepsilon \sum_{i=1}^{d} \int_{\{\bar{p}\in Q_i\}\setminus\{\bar{p}\in Q_i^\varepsilon\}} |u(x) - \bar{u}(x)|\,\mathrm{d}x\\ &\qquad \geq \varepsilon \|u - \bar{u}\|_{L^1(\Omega)} - \varepsilon \|u - \bar{u}\|_{L^\infty(\Omega)} \sum_{i=1}^{d} \int_{\{\bar{p}\in Q_i\}\setminus\{\bar{p}\in Q_i^\varepsilon\}} 1\,\mathrm{d}x,\end{aligned}$$

where we have used the L^∞-boundedness of $u - \bar{u}$ in the last step. We now use Assumption REG to estimate the remaining sum, yielding

$$\sum_{i=1}^{d} \int_{\{\bar{p}\in Q_i\}\setminus\{\bar{p}\in Q_i^\varepsilon\}} 1\,\mathrm{d}x = \operatorname{meas}\left(\bigcup_{i=1}^{d-1} \left\{ x \in \Omega : \left|\bar{p}(x) - \frac{\alpha}{2}(u_i + u_{i+1})\right| < \varepsilon \right\} \right) \leq c\varepsilon^\kappa.$$

Summarizing, we have for a constant $c > 1$ that

$$(-\bar{p}, u - \bar{u})_{L^2(\Omega)} + \alpha G'(\bar{u}; u - \bar{u}) \geq \varepsilon \|u - \bar{u}\|_{L^1(\Omega)} - c\varepsilon^{\kappa+1},$$

and hence setting

$$\varepsilon := c^{-\frac{2}{\kappa}} \|u - \bar{u}\|_{L^1(\Omega)}^{\frac{1}{\kappa}}$$

finishes the proof. □

We now have everything at hand to prove approximation error estimates.

Theorem 3.4. *Let $\bar{u}$ be a solution of (P) and assume that Assumption REG is satisfied. Furthermore, let u_γ be the solution of (P_γ) for $\gamma > 0$. Then there exists a constant $c > 0$ such that*

$$\frac{1}{\gamma}\|y_\gamma - \bar{y}\|_Y^2 + \frac{1}{\gamma}\|u_\gamma - \bar{u}\|_{L^1(\Omega)}^{1+\frac{1}{\kappa}} + \|u_\gamma - \bar{u}\|_{L^2(\Omega)}^2 \leq c\gamma^\kappa.$$

Proof. First note that G is a convex function and hence that

$$G'(\bar{u}; u_\gamma - \bar{u}) + G'(u_\gamma; \bar{u} - u_\gamma) \leq 0.$$

We thus obtain from Proposition 2.3 and Lemma 3.3 that

$$\begin{aligned} (-\bar{p}, u - \bar{u})_{L^2(\Omega)} + \alpha G'(\bar{u}; u - \bar{u}) &\geq c_A \|u_\gamma - \bar{u}\|_{L^1(\Omega)}^{1+\frac{1}{\kappa}} \quad && \forall u \in U_{\text{ad}}, \\ (-p_\gamma, u - u_\gamma)_{L^2(\Omega)} + \alpha G'(u_\gamma; u - u_\gamma) + \gamma(u_\gamma, u - u_\gamma)_{L^2(\Omega)} &\geq 0 && \forall u \in U_{\text{ad}}. \end{aligned}$$

Adding both inequalities yields

$$\begin{multline} (-\bar{p} + p_\gamma + \gamma u, u_\gamma - \bar{u})_{L^2(\Omega)} + \alpha(G'(\bar{u}; u_\gamma - \bar{u}) + G'(u_\gamma; \bar{u} - u_\gamma)) + \gamma(u_\gamma, \bar{u} - u_\gamma)_{L^2(\Omega)} \\ \geq c_A \|u_\gamma - \bar{u}\|_{L^1(\Omega)}^{1+\frac{1}{\kappa}}. \end{multline}$$

Hence, Young's inequality gives

$$\begin{aligned} \|y_\gamma - \bar{y}\|_Y^2 + c_A \|u_\gamma - \bar{u}\|_{L^1(\Omega)}^{1+\frac{1}{\kappa}} + \gamma \|u_\gamma - \bar{u}\|_{L^2(\Omega)}^2 &\leq \alpha(G'(\bar{u}; u_\gamma - \bar{u}) + G'(u_\gamma; \bar{u} - u_\gamma)) \\ &\quad + \gamma(\bar{u}, \bar{u} - u_\gamma)_{L^2(\Omega)} \\ &\leq \gamma(\bar{u}, \bar{u} - u_\gamma)_{L^2(\Omega)} \\ &\leq c\gamma \|u_\gamma - \bar{u}\|_{L^1(\Omega)} \\ &\leq \frac{c_A}{2} \|u_\gamma - \bar{u}\|_{L^1(\Omega)}^{1+\frac{1}{\kappa}} + c\gamma^{k+1}, \end{aligned}$$

from which the stated inequality follows immediately. □

4 DISCRETIZATION ERROR ESTIMATES

In practice, the exact operator K is not realizable, and a discretization $K_h : L^2(\Omega) \longrightarrow Y_h$ with finite dimensional range Y_h must be employed. Denote by $u_{\gamma,h}, y_{\gamma,h}, p_{\gamma,h}$ the solution of the discrete problem

$$(P_{\gamma,h}) \qquad \min_{u \in U_{\text{ad}}} \frac{1}{2}\|K_h u - z\|_Y^2 + \alpha G(u) + \frac{\gamma}{2}\|u\|_{L^2(\Omega)}^2.$$

If K is the solution operator of an elliptic partial differential equation and K_h its finite element discretization as in the next section, ($P_{\gamma,h}$) can be interpreted as a variational discretization [12, 13].

We assume that for all $h > 0$ the estimate

$$\|(K - K_h)u_{\gamma,h}\|_Y + \|(K^* - K_h^*)(y_{\gamma,h} - z)\|_{L^2(\Omega)} \leq \delta(h), \tag{4.1}$$

holds with a monotonically increasing function $\delta : \mathbb{R}_0^+ \to \mathbb{R}$ such that $\delta(0) = 0$. Note that this approximation condition only needs to be satisfied for the solutions to the *discretized* problem ($P_{\gamma,h}$). However, as in [23] the condition can also be replaced by a corresponding condition for the solution to the continuous problem (P_γ).

Now, we follow [23, Proposition 1.8] and estimate the discretization error for the solution to (P_γ).

Theorem 4.1. *For all $\gamma > 0$ and $h \geq 0$ there holds*

$$\|y_\gamma - y_{\gamma,h}\|_Y^2 + \gamma\|u_\gamma - u_{\gamma,h}\|_{L^2(\Omega)}^2 \leq (1 + \gamma^{-1})\delta(h)^2.$$

Proof. With $u_{\gamma,h}$ and u_γ solutions to ($P_{\gamma,h}$) and (P_γ), respectively, we have from Proposition 2.3 that

$$\begin{aligned} \left(-p_{\gamma,h} + \gamma u_{\gamma,h}, u_\gamma - u_{\gamma,h}\right)_{L^2(\Omega)} + \alpha G'(u_{\gamma,h}; u_\gamma - u_{\gamma,h}) &\geq 0, \\ \left(-p_\gamma + \gamma u_\gamma, u_{\gamma,h} - u_\gamma\right)_{L^2(\Omega)} + \alpha G'(u_\gamma; u_{\gamma,h} - u_\gamma) &\geq 0. \end{aligned}$$

Adding these two inequalities, substituting $p_{\gamma,h} = -K_h(K_h u_{\gamma,h} - z), p_\gamma = -K(Ku_\gamma - z)$, and using the convexity of G then yields

$$\begin{aligned} &\left(K_h^*(K_h u_{\gamma,h} - z) + \gamma u_{\gamma,h}, u_\gamma - u_{\gamma,h}\right) + \left(K^*(Ku_\gamma - z) + \gamma u_\gamma, u_{\gamma,h} - u_\gamma\right) \\ &\qquad\qquad \geq -\alpha\left(G'(u_{\gamma,h}; u_\gamma - u_{\gamma,h}) + G'(u_\gamma; u_{\gamma,h} - u_\gamma)\right) \geq 0. \end{aligned}$$

We thus obtain that

$$\begin{aligned} \gamma\|u_{\gamma,h} - u_\gamma\|_{L^2(\Omega)}^2 &\leq \left(K_h^*(y_{\gamma,h} - z) - K^*(y_\gamma - z), u_\gamma - u_{\gamma,h}\right) \\ &\leq \left((K_h^* - K^*)(y_{\gamma,h} - z), u_\gamma - u_{\gamma,h}\right) + \left(K^*(y_{\gamma,h} - y_\gamma), u_\gamma - u_{\gamma,h}\right). \end{aligned}$$

The rest of the proof follows similarly to the proof of [23, Proposition 1.6]. The first term on the right-hand side is estimated by the Cauchy–Schwarz inequality and the inequality (4.1) as

$$\left((K_h^* - K^*)(y_{\gamma,h} - z), u_\gamma - u_{\gamma,h}\right) \leq \frac{\gamma}{2}\|u_{\gamma,h} - u_\gamma\|_{L^2(\Omega)}^2 + \frac{1}{2\gamma}\delta(h)^2.$$

Rewriting the second term and using again the Cauchy–Schwarz inequality combined with the inequality (4.1), we obtain

$$\begin{aligned} \left(K^*(y_{\gamma,h} - y_\gamma), u_\gamma - u_{\gamma,h}\right) &= -\|y_\gamma - y_{\gamma,h}\|_Y^2 + (y_\gamma - y_{\gamma,h}, (K_h - K)u_{\gamma,h}) \\ &\leq -\frac{1}{2}\|y_\gamma - y_{\gamma,h}\|_Y^2 + \frac{1}{2}\delta(h)^2. \end{aligned}$$

Adding these two estimates, we finally arrive at

$$\frac{1}{2}\|y_\gamma - y_{\gamma,h}\|_Y^2 + \frac{\gamma}{2}\|u_\gamma - u_{\gamma,h}\|_{L^2(\Omega)}^2 \leq \left(\frac{1}{2} + \frac{1}{2\gamma}\right)\delta(h)^2. \qquad \square$$

Combining the approximation error estimate from Theorem 3.4 and the discretization error estimate from Theorem 4.1, we immediately obtain the following result.

Theorem 4.2. *If $\bar u$ satisfies Assumption REG, then*

$$\frac{1}{\gamma}\|y_{\gamma,h}-\bar y\|_Y^2+\|u_{\gamma,h}-\bar u\|_{L^2(\Omega)}^2\le c\left(\gamma^{-1}(1+\gamma^{-1})\delta(h)^2+\gamma^\kappa\right)$$

holds for all $\gamma > 0$ and $h \geq 0$.

5 ACTIVE SET METHOD FOR THE REGULARIZED PROBLEM

Let us now consider the special case where $y = Ku$ is given as the unique solution of the partial differential equation

$$\begin{cases} Ay = u & \text{in} \quad \Omega,\\ y = 0 & \text{on} \quad \partial\Omega.\end{cases} \tag{5.1}$$

with A being a second-order linear differential operator, e.g., $A = -\Delta$. In this case, the optimality conditions from Proposition 2.2 can be solved using a superlinearly convergent semi-smooth Newton method in function space; see [3, 6, 7].

We recall that (2.3) can be written as $u_\gamma \in H_\gamma(p_\gamma)$ for

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

where p_γ is the solution to the adjoint equation

$$\begin{cases} A^*p = z-y_\gamma & \text{in} \quad \Omega,\\ p = 0 & \text{on} \quad \partial\Omega,\end{cases} \tag{5.2}$$

and y_γ is the solution to (5.1) with $u = u_\gamma$. From the regularity theory for (5.2) and the general theory of semi-smooth Newton methods in function space [22], we deduce that the superposition operator H_γ is Newton differentiable with

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

A Newton step for the solution of (P_γ) can therefore be formulated as

$$\begin{pmatrix} -\mathrm{Id} & A & 0\\ 0 & \mathrm{Id} & A^*\\ 0 & A & -D_NH_\gamma(p^k)\end{pmatrix}\begin{pmatrix} u^{k+1}-u^k\\ y^{k+1}-y^k\\ p^{k+1}-p^k\end{pmatrix}=-\begin{pmatrix} Ay^k-u^k\\ A^*p^k+y^k-z\\ Ay^k-H_\gamma(p^k)\end{pmatrix} \tag{5.3}$$

In [3], this was reduced to a symmetric system in (y, p). Here, we instead consider an equivalent primal active set formulation that has proven to be more robust for small values of γ and h. In a slight abuse of notation, we introduce

$$Q_i^k := \left\{x \in \Omega : p^k(x) \in Q_i^\gamma\right\}, \qquad 1 \le i \le d,$$

and similarly for $Q_{i,i+1}^k$. We denote by $\mathbb{1}_C$ the characteristic function of the set C, i.e., $\mathbb{1}_C(x) = 1$ if $x \in C$ and 0 else. The following algorithm is an extension of the one proposed in [20] for $G(u) = \|u\|_{L^1(\Omega)}$.

Algorithm 1. Choose initial data u^0, p^0 and parameters α, γ, set $k = 0$ and compute the sets Q_i^0 for $1 \le i \le d$ and $Q_{i,i+1}^0$ for $1 \le i < d$.

1. Solve for $(u^{k+1}, y^{k+1}, p^{k+1}, \lambda^{k+1})$ satisfying

$$\text{(5.4a)} \qquad \begin{cases} Ay^{k+1} - u^{k+1} = 0, \\ A^*p^{k+1} + y^{k+1} - z = 0, \\ -p^{k+1} + \gamma u^{k+1} + \alpha\lambda^{k+1} = 0, \end{cases}$$

$$\text{(5.4b)} \quad \left(1 - \sum_{i=1}^{d} \mathbb{1}_{Q_i^k}\right)\lambda^{k+1} + \left(1 - \sum_{i=1}^{d-1} \mathbb{1}_{Q_{i,i+1}^k}\right)u^{k+1} = \sum_{i=1}^{d} \mathbb{1}_{Q_i^k} u_i + \frac{1}{2}\sum_{i=1}^{d-1} \mathbb{1}_{Q_{i,i+1}^k}(u_i + u_{i+1}),$$

2. Compute the sets Q_i^{k+1} for $1 \le i \le d$ and $Q_{i,i+1}^{k+1}$ for $1 \le i < d$.

3. If $Q_i^k = Q_i^{k+1}$ for $1 \le i \le d$ and $Q_{i,i+1}^k = Q_{i,i+1}^{k+1}$ for $1 \le i < d$, then go to step 4. Otherwise set $k = k + 1$ and go to step 2.

4. STOP: u^{k+1} is a solution of (P_γ).

The stopping criterion yields solutions of (P_γ).

Lemma 5.1. *If*

$$\begin{aligned} Q_i^k &= Q_i^{k+1} \quad 1 \le i \le d, \\ Q_{i,i+1}^k &= Q_{i,i+1}^{k+1} \quad 1 \le i < d, \end{aligned}$$

then the solution (u^{k+1}, p^{k+1}) computed from (5.4) satisfy (2.3). In particular, u^{k+1} is a solution to (P_γ).

Proof. Since for fixed Q_i^k and $Q_{i,i+1}^k$ the solution of (5.4) is unique, we have $(u^k, y^k, p^k) = (u^{k+1}, y^{k+1}, p^{k+1})$. Inserting this into (5.4b) and comparing with (2.3) yields the claim. □

We now show that Algorithm 1 coincides with a semi-smooth Newton, which implies locally superlinear convergence.

Theorem 5.2. *The active set step* (5.4) *is equivalent to the semi-smooth Newton step* (5.3).

Proof. Clearly, the first two equations of (5.3) are equivalent to the first two equation of (5.4a). It therefore remains to consider the last equation, which is given by

$$A(y^{k+1} - y^k) - D_N H_\gamma(p^k)(p^{k+1} - p^k) = -Ay^k - H_\gamma(p^k). \tag{5.5}$$

Let us define the function

$$\lambda^{k+1}(x) := \begin{cases} -\frac{1}{\alpha}\left(-p^{k+1}(x) + \gamma u^{k+1}\right) & \text{if } x \in Q_i^k, \\ \frac{1}{2}(u_i + u_{i+1}) & \text{if } x \in Q_{i,i+1}^k. \end{cases}$$

We now make a case distinction pointwise almost everywhere.

(i) If $x \in Q_i^k$, (5.5) reduces to $[Ay^{k+1}](x) = u_i$, and from the first line of (5.3) we obtain $u^{k+1}(x) = u_i$.

(ii) If $x \in Q_{i,i+1}^k$, (5.5) shows that

$$\gamma u^{k+1}(x) - p^{k+1}(x) + \frac{\alpha}{2}(u_i + u_{i+1}) = \gamma u^{k+1}(x) - p^{k+1}(x) + \alpha\lambda^{k+1}(x) = 0.$$

Hence the third row of (5.3) is equivalent to (5.4b). In both cases, we obtain from the definition of λ^{k+1} that

$$-p^{k+1} + \gamma u^{k+1} + \alpha\lambda^{k+1} = 0,$$

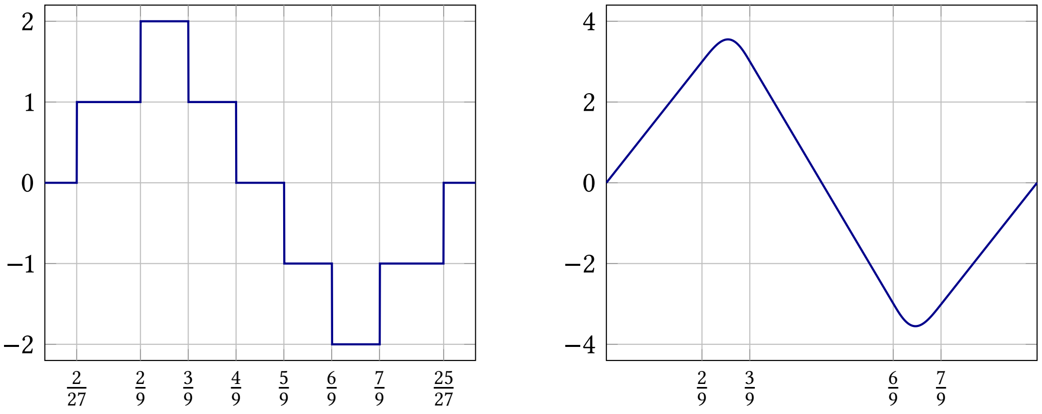
which finally gives (5.4a) and therefore the claimed equivalence. □

6 NUMERICAL RESULTS

In this section we present some numerical results and convergence rates. Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain and K be the operator mapping u to the weak solution y of

$$\begin{cases} -\Delta y = u & \text{in} \quad \Omega \\ \quad y = 0 & \text{on} \quad \partial\Omega. \end{cases} \tag{6.1}$$

The operator K_h is correspondingly defined via the Galerkin approximation of (6.1) using linear finite elements on a triangulation of Ω, which is chosen in such a way that the approximation condition (4.1) is satisfied; see [23]. For the multibang penalty, we take $(u_1, \dots, u_5) = (-2, -1, 0, 1, 2)$ and $\alpha = 2$. We implemented Algorithm 1 in Python using DOLFIN [16, 17], which is part of the open-source computing platform FEniCS [1, 15]. The linear system (5.4) arising from the active set step is solved using the sparse direct solver `spsolve` from SciPy. The code used to obtain the following results can be downloaded from https://github.com/clason/multibangestimates.



(a) optimal control $\bar{u}$

(b) optimal adjoint state $\bar{p}$

Figure 1: constructed one-dimensional example

One-dimensional example We first consider $\Omega = (0, 1)$ and define

$$
\begin{aligned}
\bar{p}(x) &:= \left(\tfrac{27}{2}x\right) \mathbb{1}_{[0,\frac{2}{9})}(x) \\
&\quad + \left(-72 + \tfrac{3123x}{2} - 13122x^2 + 54675x^3 - 111537x^4 + \tfrac{177147}{2}x^5\right) \mathbb{1}_{[\frac{2}{9},\frac{3}{9})}(x) \\
&\quad + (9 - 18x)\, \mathbb{1}_{[\frac{3}{9},\frac{6}{9})}(x) \\
&\quad + \left(-20079 + 136062x - 367416x^2 + 494262x^3 - \tfrac{662661}{2}x^4 + \tfrac{177147}{2}x^5\right) \mathbb{1}_{[\frac{6}{9},\frac{7}{9})}(x) \\
&\quad + \left(-\tfrac{27}{2} + \tfrac{27}{2}x\right) \mathbb{1}_{[\frac{7}{9},1]}(x), \\
\bar{u}(x) &:= \mathbb{1}_{[\frac{2}{27},\frac{2}{9})}(x) + 2\mathbb{1}_{[\frac{2}{9},\frac{3}{9})}(x) + \mathbb{1}_{[\frac{3}{9},\frac{4}{9})}(x) - \mathbb{1}_{[\frac{5}{9},\frac{6}{9})}(x) - 2\mathbb{1}_{[\frac{6}{9},\frac{7}{9})}(x) - \mathbb{1}_{[\frac{7}{9},\frac{25}{27})}(x), \\
\bar{y}(x) &:= \sin(2\pi x) \\
e_\Omega &:= -\Delta\bar{y} - \bar{u}, \\
z &:= -Ke_\Omega - \Delta\bar{p} + \bar{y},
\end{aligned}
$$

see Figure 1. Note that $\bar{p}, \bar{y} \in C^2(\overline{\Omega})$, and that $\bar{u}$ and $\bar{p}$ satisfy the optimality conditions in Proposition 2.1. Hence, $(\bar{u}, \bar{p})$ are a solution to (P). From Theorem 3.2 we further deduce that Assumption REG is satisfied with $\kappa = 1$.

We now compute the solution of $(P_{\gamma,h})$ for different values of h, where Ω is divided into equidistant elements with mesh size h. From Theorem 3.4 we expect that the numerical convergence rate

$$
\kappa_\gamma := \frac{1}{\log(2)} \log\left(\frac{\|u_{\gamma/2,h} - \bar{u}\|^2_{L^2(\Omega)}}{\|u_{\gamma,h} - \bar{u}\|^2_{L^2(\Omega)}}\right)
$$

satisfies $\kappa_\gamma \geq \kappa = 1$. Note that for $d = 2$, it is known that Assumption REG is not only sufficient for convergence rates similar to Theorem 3.4 but also necessary for high convergence rates; see [24]. Hence, we expect that $\kappa_\gamma \approx 1$, which can be observed from Table 1a and Figure 2a. In addition, we observe that for small γ the discretization error dominates, which is an expected saturation effect; see Theorem 4.2.

Table 1: computed numerical order of convergence for different h

(a) one-dimensional example

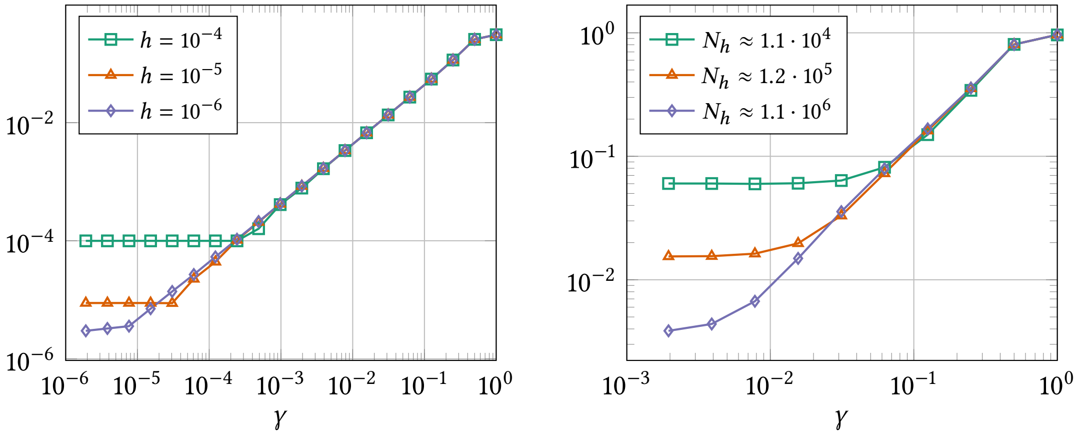
	κ_γ		
$\gamma \setminus h$	10^{-4}	10^{-5}	10^{-6}
2^{-1}	0.2621	0.2621	0.2621
2^{-2}	1.1608	1.1608	1.1608
2^{-3}	1.0735	1.0735	1.0735
2^{-4}	1.0143	1.0142	1.0141
2^{-5}	1.0036	1.0033	1.0033
2^{-6}	1.0028	1.0008	1.0007
2^{-7}	1.0003	1.0003	1.0001
2^{-8}	1.0211	1.0004	0.9998
2^{-9}	1.0897	1.0016	0.9994
2^{-10}	0.9295	1.0038	0.9989
2^{-11}	1.3559	1.0200	0.9978
2^{-12}	0.6828	1.0049	0.9954
2^{-13}	0.0	1.2315	0.9923
2^{-14}	0.0	0.9592	0.9917
2^{-15}	0.0	1.3658	0.9534
2^{-16}	0.0	−0.0096	0.9701
2^{-17}	0.0	0.0	0.9814
2^{-18}	0.0	0.0	0.1308
2^{-19}	0.0	0.0	0.1312

(b) two-dimensional example

	κ_γ		
$\gamma \setminus N_h$	$1.1 \cdot 10^4$	$1.2 \cdot 10^5$	$1.1 \cdot 10^6$
2^{-1}	0.2550	0.2629	0.2634
2^{-2}	1.2370	1.1808	1.1712
2^{-3}	1.1908	1.1377	1.0985
2^{-4}	0.8784	1.1503	1.0750
2^{-5}	0.3564	1.1367	1.1517
2^{-6}	0.0733	0.7440	1.2583
2^{-7}	0.0167	0.2779	1.1530
2^{-8}	−0.0094	0.0669	0.6110
2^{-9}	−0.0010	0.0072	0.1839

Two-dimensional example We now consider the circular domain $\Omega := \{x \in \mathbb{R}^2 : \|x\|_2 < 1\}$ and define the following radially symmetric functions, where we set $r = r(x) := \|x\|_2$ for conciseness.

$$\begin{aligned}
\bar{p}(r) := {} & \left(\tfrac{59049}{4} r^3 - \tfrac{531441}{2} r^4 + \tfrac{43046721}{32} r^5\right) \mathbb{1}_{[0, \frac{2}{27})}(r) + \left(\tfrac{27}{2} r\right) \mathbb{1}_{[\frac{2}{27}, \frac{2}{9})}(r) \\
& + \left(-72 + \tfrac{3123}{2} r - 13122 r^2 + 54675 r^3 - 111537 r^4 + \tfrac{177147}{2} r^5\right) \mathbb{1}_{[\frac{2}{9}, \frac{3}{9})}(r) \\
& + (9 - 18r)\, \mathbb{1}_{[\frac{3}{9}, \frac{6}{9})}(r) \\
& + \left(-20079 + 136062 r - 367416 r^2 + 494262 r^3 - \tfrac{662661}{2} r^4 + \tfrac{177147}{2} r^5\right) \mathbb{1}_{[\frac{6}{9}, \frac{7}{9})}(r) \\
& + \left(-\tfrac{27}{2} + \tfrac{27}{2} r\right) \mathbb{1}_{[\frac{7}{9}, 1]}(r), \\
\bar{u}(r) := {} & \mathbb{1}_{[\frac{2}{27}, \frac{2}{9})}(r) + 2\mathbb{1}_{[\frac{2}{9}, \frac{3}{9})}(r) + \mathbb{1}_{[\frac{3}{9}, \frac{4}{9})}(r) - \mathbb{1}_{[\frac{5}{9}, \frac{6}{9})}(r) - 2\mathbb{1}_{[\frac{6}{9}, \frac{7}{9})}(r) - \mathbb{1}_{[\frac{7}{9}, \frac{25}{27})}(r), \\
\bar{y}(r) := {} & \sin(2\pi r^2), \\
e_\Omega := {} & -\Delta \bar{y} - \bar{u}, \\
z := {} & -K e_\Omega - \Delta \bar{p} + \bar{y}.
\end{aligned}$$



(a) one-dimensional example

(b) two-dimensional example

Figure 2: discretization error $\|u_{\gamma,h} - \bar{u}\|^2_{L^2(\Omega)}$ for different γ and h

Note that these functions are similar to the functions used in the one-dimensional example. Again one can check that $(\bar{u}, \bar{p})$ satisfy the optimality conditions from Proposition 2.1. Using again Theorem 3.2, we obtain that Assumption REG is satisfied with $\kappa = 1$.

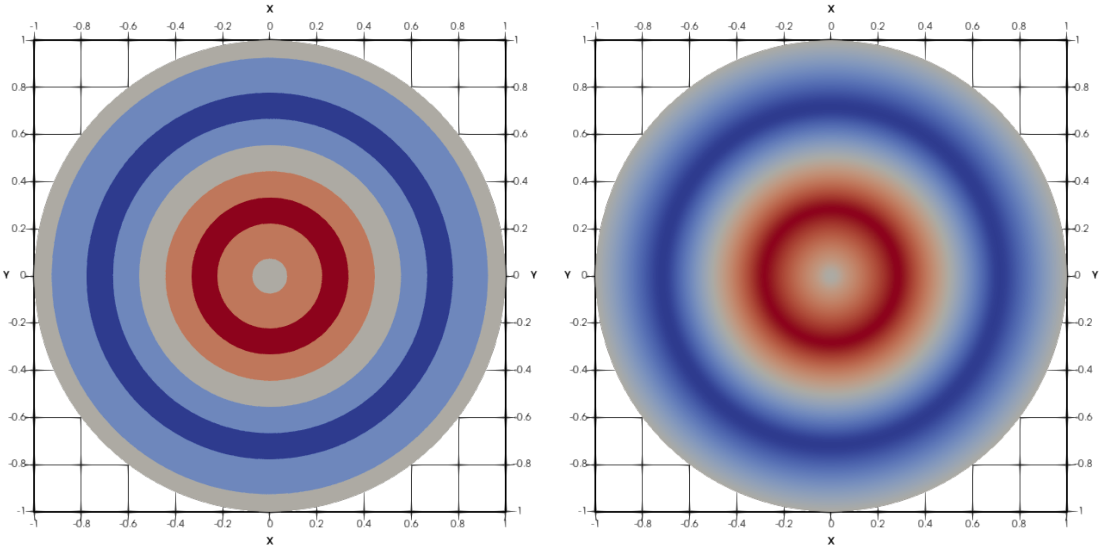
We now use a regular triangulation of the domain Ω, whose degrees of freedom are denoted with N_h. A plot of the computed solution can be found in Figure 3. The obtained convergence rates can be seen in Table 1b and Figure 2b and show a similar behavior as in the one-dimensional case, including the saturation effect for small γ.

7 CONCLUSIONS

For optimal control problems with a convex penalty promoting minimizers that pointwise almost everywhere take on values from a given discrete set, Moreau–Yosida approximation allows the solution by a superlinearly convergent semi-smooth Newton method. On a structural assumption on the behavior of the adjoint state near singular sets, convergence rates as the approximation parameter $\gamma \to 0$ can be derived. The same assumption also yields discretization error estimates for fixed $\gamma > 0$. Numerical experiments corroborate the predicted rate.

This work can be extended in a number of directions. First, an active set condition similar to Assumption REG was derived in [19] for the approximation of bang-bang control of a semilinear equation and could be adapted to the multibang control setting. Of particular interest would be the extension to problems where the control enters into the principal part of an elliptic equation as in the case of topology optimization problems [5, 7].

On the other hand, the applicability of the multibang penalty G to the regularization of inverse problems was demonstrated in [3]. There, a condition related to Assumption REG was used to derive strong convergence as $\alpha \to 0$, albeit without rates; and a natural question is whether the more quantitative Assumption REG would allow obtaining such rates at least in $L^2(\Omega)$. Finally, combined regularization, approximation, and discretization estimates for the convergence $(\alpha, \gamma, h) \to 0$ would be highly useful.



(a) optimal control $u_{\gamma,h}$

(b) optimal adjoint state $p_{\gamma,h}$

Figure 3: computed solution for the two-dimensional problem for $\gamma = 2^{-9}$ and $N_h = 9.0 \cdot 10^5$

ACKNOWLEDGMENT

This work was funded by the German Research Foundation (DFG) under grants Cl 487/1-1 and Wa 3626/1-1.

REFERENCES

[1] Alnæs, Blechta, Hake, Johansson, Kehlet, Logg, Richardson, Ring, Rognes & Wells, The FEniCS Project Version 1.5, *Archive of Numerical Software* 3 (2015), 9–23, doi: 10.11588/ans.2015.100.20553.

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

[3] Clason & Do, Convex regularization of discrete-valued inverse problems, in: *New Trends in Parameter Identification for Mathematical Models*, Springer, 2018, doi: 10.1007/978-3-319-70824-9_2.

[4] 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.

[5] Clason, Kruse & Kunisch, Total variation regularization of multi-material topology optimization, *ESAIM: Mathematical Modelling and Numerical Analysis* Forthcoming Article (2018), doi: 10.1051/m2an/2017061.

[6] 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.

[7] 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.

[8] Clason, Tameling & Wirth, Vector-valued multibang control of differential equations, *SIAM Journal on Control and Optimization* to appear (2018), arxiv: 1611.07853.

[9] Deckelnick & Hinze, A note on the approximation of elliptic control problems with bang-bang controls, *Comput. Optim. Appl.* 51 (2012), 931–939, doi: 10.1007/s10589-010-9365-z.

[10] Hintermüller & Hinze, Moreau-Yosida regularization in state constrained elliptic control problems: error estimates and parameter adjustment, *SIAM J. Numer. Anal.* 47 (2009), 1666–1683, doi: 10.1137/080718735.

[11] Hintermüller, Schiela & Wollner, The length of the primal-dual path in Moreau-Yosida-based path-following methods for state constrained optimal control, *SIAM J. Optim.* 24 (2014), 108–126, doi: 10.1137/120866762.

[12] Hinze, A Variational Discretization Concept in Control Constrained Optimization: The Linear-Quadratic Case, *Computational Optimization and Applications* 30 (2005), 45–61, doi: 10.1007/s10589-005-4559-5.

[13] Hinze & Matthes, A note on variational discretization of elliptic Neumann boundary control, *Control Cybernet.* 38 (2009), 577–591.

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

[15] Logg, Mardal, Wells, et al., *Automated Solution of Differential Equations by the Finite Element Method*, Springer, 2012, doi: 10.1007/978-3-642-23099-8.

[16] Logg & Wells, DOLFIN: Automated Finite Element Computing, *ACM Transactions on Mathematical Software* 37 (2010), doi: 10.1145/1731022.1731030.

[17] Logg, Wells & Hake, DOLFIN: a C++/Python Finite Element Library, in: *Automated Solution of Differential Equations by the Finite Element Method, Volume 84 of Lecture Notes in Computational Science and Engineering*, Springer, 2012, chap. 10, doi: 10.1007/978-3-642-23099-8_10.

[18] Pörner & Wachsmuth, An iterative Bregman regularization method for optimal control problems with inequality constraints, *Optimization* 65 (2016), 2195–2215, doi: 10.1080/02331934.2016.1238082.

[19] Pörner & Wachsmuth, Tikhonov regularization of optimal control problems governed by semi-linear partial differential equations, *Mathematical Control & Related Fields* 8 (2018), 315–335, doi: 10.3934/mcrf.2018013.

[20] Stadler, Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices, *Comput. Optim. Appl.* 44 (2009), 159–181, doi: 10.1007/s10589-007-9150-9.

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

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

[23] Wachsmuth, Adaptive regularization and discretization of bang-bang optimal control problems, *Electron. Trans. Numer. Anal.* 40 (2013), 249–267, url: http://etna.mcs.kent.edu/volumes/2011-2020/vol40/abstract.php?vol=40&pages=249-267.

[24] Wachsmuth & Wachsmuth, Necessary conditions for convergence rates of regularizations of optimal control problems, in: *System modeling and optimization*, Springer, Heidelberg, 2013, 145–154, doi: 10.1007/978-3-642-36062-6_15.

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

[26] 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.