SpezialForschungsBereich F 32

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

$L^p(\Omega)$- Optimization with $p \in [0; 1)$

K. Ito K. Kunisch

SFB-Report No. 2012-019 October 2012

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the
Austrian Science Fund (FWF)

SFB sponsors:

- **Austrian Science Fund** (FWF)
- **University of Graz**
- **Graz University of Technology**
- **Medical University of Graz**
- **Government of Styria**
- **City of Graz**

Das Land Steiermark

Stadt GRAZ Wissenschaft

$L^p(\Omega)-$Optimization with $p \in [0, 1)$

Kazufumi Ito[1] Karl Kunisch[2]

Graz, October 2012

[1]Department of Mathematics, North Carolina State University, Raleigh, North Carolina, 27695-8205, USA; research partially supported by the Army Research Office under DAAD19-02-1-0394

[2]Institut für Mathematik, Karl-Franzens-Universität Graz, A-8010 Graz, Austria, supported in part by the Fonds zur Förderung der wissenschaftlichen Forschung under SFB 32, Mathematical Optimization and Applications in the Biomedical Sciences".

Abstract L^p optimization with $p \in [0,1)$ is investigated. The difficulty of natural lack of weak lower-semicontinuity is addressed by introducing appropriately chosen regularization terms. Existence results and necessary optimality conditions are obtained, and convergence of a monotone scheme is proved. Special attention is given to the particular case of optimal control problems in L^0. A maximum principle is derived and existence of controls, in some cases relaxed controls, is proved, and an estimate on the consequences of relaxation are estimated.

Keywords L^0 minimization, optimal control, bang-bang control, sparsity optimization, maximum principle, non-smooth optimization, primal-dual active set method.

MSC Classification 49J20, 49J30, 49K20, 49J52.

1 Introduction

Quadratic expressions of Hilbert-space norms have been commonly used to model control costs in optimal control or to regularize ill-posed inverse problems. The reasons for this choice include their statistical interpretation and the ease of differentiation. The use of higher order polynomial powers, corresponding to L^p-norms, $p > 1$, were motivated, for instance, in optimal control of semi-linear partial differential equations to guarantee appropriate a-priori bounds. More recently the use of the ℓ^1 and L^1-norms has been recognized as a useful tool for enhancing sparsity in data-management problems and in optimal control. Since there is already a vast literature on these topics, we can only quote selected papers [1, 9, 7, 15] and the references there for imaging and [4, 6, 11] for optimal control. In robust statistics the use L^1-type functionals has a long standing history.

The question naturally arises of choosing the exponent $p < 1$ and letting it attain the value 0. It will be addressed for L^p, $p \in [0,1)$, in this paper. It continues our research from [14] where we treated sequence spaces ℓ^p with $p \in [0,1)$. The case of sequence spaces ℓ^p with $p \in (0,1)$ was also considered in [2, 10, 12, 17, 18], for example.

Let us briefly describe the contents of the paper. In Section 2 the functional analytical preliminaries are presented. The metric for L^0 is fixed to be the Ekeland metric. Its use in the cost functional implies a volume constraint. It is thus different from the commonly chosen metric of convergence in measure. Our choice of metric for L^0 corresponds to Donoho's counting norm introduced for ℓ^0, [9].

To establish existence for variational problems involving L^p functionals with $p \in [0,1)$, one cannot rely on weak lower semi-continuity properties. In the case of ℓ^p one can appropriately transform the problem and use the fact that the duality mapping from $\ell^q \to \ell^{q'}$, with $\frac{1}{p} + \frac{1}{p'} = 1$, is weakly sequentially continuous, [14, 18]. The lack of weak lower semi-continuity in the L^p-case, $p \in [0,1)$ presents a major difficulty which we propose to meet with appropriately chosen regularization terms. They depend on the structure of the problem under consideration. We shall distinguish between optimal control problems which themselves depend on the state- and control variables, and where the adjoint state determines the structure of the optimal control, and "general" problems which do not allow such a splitting and adjoint states. In the first case the regularization will be in L^2 in the latter case in H^1.

Section 3 and 4 are dedicated to optimal control problems with L^0-sparsity enhancing functions. The approach we take is that we first assume existence of optimum controls and derive a maximum principle that they must satisfy. In a second step, existence is addressed for a restricted class of problems with linear dynamics. For this purpose we analyze the optimality system by monotone operator theory techniques. Under a condition to be specified below, existence is obtained. If this condition is not fulfilled one has to pass to the maximal monotone extension of the negative feedback operator that describes the optimal control as a function of the adjoint state. This corresponds to obtaining a relaxed control. We can obtain an estimate on the error in the cost corresponding to the relaxed control and the infimum of the optimal control problem. This approach is carried out from problems without constraints on the controls in Section 3, where also the first step towards numerical realization on the basis of a primal dual active set strategy are proposed. In Section 4 the case of bilateral constraints on the control is investigated. This is of particular interest since the optimal controls, except at two critical values of the adjoint variable, assume only three values: the upper and lower bounds, and zero, i.e., they are switching controls of bang-bang-bang type.

In Section 5 we consider existence for a general class of optimization problems involving L^0 type regularization terms. It includes the optimal control problems but it is a wider class since it does not require the regularizing effect that is present in the control-to-solution mapping of optimal control problems.

In Section 6 we first consider a general class of problems with L^p, $p \in (0,1)$, regularization. The case $p = \frac{1}{2}$ is of special importance, since it provides the best fit to the heavy tailed shape of the true probability density function in image denoising, [13]. Subsequently we analyze convergence of a monotone scheme to solve the L^p-problems iteratively. A brief outlook concludes the paper.

2 Preliminaries

Let $(\Omega, \mathcal{B}, \mu)$ denote a finite measure space, where Ω is a domain in $\mathbb{R}^n$, $\mathcal{B}$ a $\sigma-$ algebra of subsets of Ω, and μ a measure with domain $\mathcal{B}$. For $0 < p < 1$, let $L^p(\Omega)$ be the vector space of measurable functions f for which

$$N_p(f) = \int_\Omega |f(x)|^p \, d\mu < \infty.$$

Since $(a+b)^p \le a^p + b^p$ is valid for any $a \ge 0$ and $b \ge 0$, N_p is sub-additive, i.e.

$$N_p(f+g) \le N_p(f) + N_p(g) \tag{2.1}$$

holds. The function

$$d_p(f,g) = N_p(f-g)$$

defines a metric on $L^p(\Omega)$. Endowed with this metric $L^p(\Omega)$ is complete. Clearly $|N_p(f)|$ is not positively homogenous and hence it is not sublinear. Moreover $N_p(f)^{\frac{1}{p}}$ is a quasi-norm, but not a norm, since the triangle inequality does not hold. In fact, we have $N_p(f+g)^{\frac{1}{p}} \le 2^{\frac{1}{p}-1}(N_p(f)^{\frac{1}{p}} + N_p(g)^{\frac{1}{p}})$. It can also be noted that the topological dual of $L^p(\Omega)$ consists only of the zero-functional for any $0 < p < 1$, [8].

For $p = 0$ one can define $L^0(\Omega)$ as the space of measurable function with Ekeland distance

$$d_0(f,g) = \text{meas}\{x \in \Omega : f(x) \neq g(x)\}.$$

This makes $L^0(\Omega)$ a complete metric space. We also define

$$N_0(f) = \text{meas}\{x \in \Omega : f(x) \neq 0\},$$

and note that $N_0(f) = d_0(f,0)$ and that N_0 is subadditive, i.e.

$$N_0(f+g) \leq N_0(f) + N_0(g). \tag{2.2}$$

Since it is also positively homogenous, N_0 is sublinear. If $(\Omega, \mathcal{B}, \mu)$ is a discrete measure space then N_0 coincides with Donoho's discrete counting measure.

For functions $f \in L^p(\Omega)$ with $0 < p < 1$ we have

$$N_p(f) = \int_{f<1} |f|^p \, dx + \int_{f\geq 1} |f|^p \, dx,$$

where the first family of integrals on the right hand side increases and the second family of integrals decreases as $p \to 0^+$. Hence by Lebegue's monotone convergence theorem

$$\lim_{p\to 0^+} N_p(f) = N_0(f) = \int_\Omega |f(x)|^0 \, dx, \tag{2.3}$$

where we set $0^0 = 0$.

Frequently $L^0(\Omega)$ is endowed with the topology of convergence in measure. This topology can be defined by any metric of the form

$$\tilde{d}_0(f,g) = \int_\Omega \varphi(|f(x) - g(x)|) \, d\mu(x),$$

where φ is any bounded, concave, continuous, non-decreasing function on $[0,\infty)$ with $\varphi(0) = 0$ and $\varphi(t) > 0$ for $t > 0$, [16], as for instance $\varphi(t) = \frac{t}{\varepsilon + t}$ or $\varphi(t) = \min(\frac{t}{\varepsilon}, 1)$, for any $\varepsilon > 0$. For these metrics, however, we do not have (2.3).

Let us further compare the d_0 and the $\tilde{d}_0$ metrics. If if $f_n \to f$ a.e. in Ω, then by Lebesgue's bounded convergence theorem $\lim_{n\to\infty} \tilde{d}(f_n, f) = 0$. This is not the case for the d_0-metric. In fact, the sequence of constant functions $f_n = \{\frac{1}{n}\}$ converges to the zero function, but $d_0(f_n, 0) = \text{meas}(\Omega)$ for all n. But d_0 has the following lower semi-continuity property:

$$N_0(u) \leq \liminf_{n\to\infty} N_0(u_n), \qquad \text{for } u_n \to u \text{ a.e. in } \Omega, \tag{2.4}$$

which follows from Fatou's lemma. Moreover, on the set

$$S = \{f : \Omega \to \mathbb{R} : \text{ measurable}, |f(x)| \leq 1 \; a.e. \text{ in } \Omega\},$$

we have

$$N_0(f) \geq N_p(f) \geq N_q(f) \quad \text{for } 0 < p < q \leq 1,$$

with strict inequalities holding if $f \neq 0$ and $f \neq 1$ a.e. This property does not hold for d_0 replaced by $\tilde{d}_0$.

We note that N_0 is not convex which can be seen by setting $f = 0$ and $g \neq 0$. If f and g are different from 0 a.e., then for any $\lambda \in [0, 1]$ we have

$$N_0(\lambda f + (1 - \lambda)g) \leq \lambda N_0(f) + (1 - \lambda)N_0(g).$$

3 $L^0(\Omega)$ optimal control

3.1 Necessary optimality

Due to difficulties associated to the N_0-functional, in particular the lack of weak lower semi-continuity properties, we start by deriving necessary optimality conditions. Then we analyze these necessary conditions.

We first consider a class of optimization problems associated to optimal control problems with cost functionals which are not necessarily convex, and in particular will be applicable to functionals involving N_0 as introduced in the previous section.

We denote by X a Banach space that is densely and compactly embedded in to $L^2(\Omega)$. Here Ω is a bounded subset of $\mathbb{R}^n$ which describes the (time)-space domain of the control system and is endowed with the Lebesgue measure. The state variable and control variable variables are denoted by x and u respectively.

We consider the constrained minimization problem

$$\min \quad \int_\Omega \big(\ell(\omega, x(\omega)) + h(u(\omega))\big)\, d\omega \tag{3.1}$$

subject to the equality constraint

$$E\, x + f(\cdot, x, u) = 0 \quad \text{in } X^* \tag{3.2}$$

over

$$u \in U_{ad} = \{u \in L^2(\Omega) : u(\omega) \in U \;\; a.e\}. \tag{3.3}$$

Here U is a closed convex subset of $\mathbb{R}$ and $E \in \mathcal{L}(X, X^*)$, with E^* the dual space to E. Further $f \in C^2(\mathbb{R}^3, \mathbb{R})$, $\ell \in C^2(\mathbb{R}^2, \mathbb{R})$ and $h \in C(\mathbb{R}, \mathbb{R})$. The mappings f, ℓ, h give rise to substitution operators which are denoted by the same symbols and are supposed to satisfy:

$$\begin{cases} (x,u) \in X \times U_{ad} \to f(\cdot, x, u) \in L^2(\Omega), \\ x \in X \to \ell(\cdot, x) \in L^1(\Omega), \\ u \in U_{ad} \to h(u) \in L^1(\Omega). \end{cases} \tag{3.4}$$

Throughout it is assumed that (3.2) admits a solution $x = x(u) \in X$ for any $u \in L^2(\Omega)$.

Unless otherwise specified we also assume the existence of a solution $\bar{u}$ for (3.1) – (3.3) with associated state $\bar{x} = x(\bar{u})$.

To derive a necessary condition for this class of (nonconvex) problems we use a maximum principle approach and define the Hamiltonian $\mathcal{H} : \mathbb{R}^4 \to \mathbb{R}$ by

$$\mathcal{H}(\omega, x, u, p) = \ell(\omega, x) + h(u) + p \cdot f(\omega, x, u).$$

We require in addition that the following substitution operators are well-defined:

$$\begin{cases} (x,u) \in X \times U_{ad} \to f_x(\cdot, x, u) \in \mathcal{L}(X, X^*), \\ x \in X \to \ell_x(\cdot, x) \in \mathcal{L}(X, L^1(\Omega) \cap X^*). \end{cases} \tag{3.5}$$

With (3.5) holding the following adjoint equation is well-defined:

$$(E + f_x(\cdot, \bar{x}, \bar{u}))^* p + \ell_x(\cdot, \bar{x}) = 0. \tag{3.6}$$

It will be assumed that (3.6) admits a unique solution $p \in X$.

For arbitrary $s \in \Omega$, we shall utilize needle perturbations of the optimal control defined by

$$v(\omega) = \begin{cases} u & \text{on } B(s,\delta) = \{\omega : |\omega - s| < \delta\} \\ \bar{u}(\omega) & \text{otherwise,} \end{cases} \tag{3.7}$$

where $u \in U$ is constant and $\delta > 0$ is sufficiently small so that $B(s,\delta) \subset \Omega$. With our choice of needle perturbations we do not aim for the widest possible generality. In this respect we refer to more elaborate needle perturbation techniques as used for example in [5].

We denote by $x = x(v)$ the solution corresponding solution of (3.2). The following additional properties for the optimal state $\bar{x}$ and each perturbed state v will be needed:

$$\begin{cases} |x(v) - \bar{x}|^2_{L^2(\Omega)} = o\,(\text{meas}(B(s,\delta))) \\ \int_\Omega \big(\, \ell(\cdot, x(v)) - l(\cdot, \bar{x}) - \ell_x(\cdot, \bar{x})(x(v) - \bar{x})\big) d\omega \, = O(|x(v) - \bar{x}|^2_{L^2}) \\ \langle f(\cdot, x(v), v) - f(\cdot, \bar{x}, v) - f_x(\cdot, \bar{x}, v)(x(v) - \bar{x}), p\rangle_{X^*,X} \, = O(|x(v) - \bar{x}|^2) \\ (f_x(\cdot, \bar{x}, v) - f_x(\cdot, \bar{x}, \bar{u}))p \in L^2(\Omega), \text{ for all } v \text{ defined in (3.7).} \end{cases} \tag{3.8}$$

Remark 3.1. The first assumption in (3.8) is well-established in the context of ordinary differential equation. In the case of elliptic systems the following considerations can be used to establish this condition. We assume that $f(\cdot, x, u) = Bu$ with $B \in \mathcal{L}(U, X) \cap \mathcal{L}(L^1(\Omega))$ and that there exists $\omega > 0$ such that

$$\omega\, |x_1 - x_2|^2_X \le \langle E(x_1 - x_2), x_1 - x_2\rangle_{X^*,X}$$

for all $x_1,\, x_2 \in X$. Then for every $v \in L^2(\Omega)$ there exists a unique solution $x = x(v)$ to $Ex + Bu = 0$ and we have

$$\langle E(x-\bar{x}), x-\bar{x}\rangle_{X^*,X} = |(B(v-\bar{u}), x-\bar{x})| \le |B|_{\mathcal{L}(L^1(\Omega))}|x-\bar{x}|_{L^\infty(\Omega)}|v-\bar{u}|_{L^1(\Omega)}.$$

Let us further assume that

$$|x(v_\delta) - \bar{x}|_{L^\infty(\Omega)} \to 0 \text{ as } \delta \to 0^+,$$

where v_δ is defined according to (3.7). Then by the above estimates

$$|x(v_\delta) - \bar{x}|^2_X \le o(meas(B(s,\delta))),$$

which implies the first estimate in (3.8). For E an elliptic operator with sufficiently smooth coefficients we have

$$|x - \bar{x}|_{H^2(\Omega)} \le M\, |v - \bar{u}|_{L^2(\Omega)} \sim O(\sqrt{meas(B(s,\delta))}),$$

and hence $|x - \bar{x}|_{L^\infty(\Omega)} \sim O(\sqrt{meas(B(s,\delta))})$ in dimension 2 or 3.

Theorem 3.1. *Suppose* $(\bar{x}, \bar{u}) \in X \times U_{ad}$ *is optimal for problem* (3.1) *, that* $p \in X$ *satisfies the adjoint equation* (3.6) *and that* (3.4)*,* (3.5)*, and* (3.8) *hold. Then we have the necessary optimality condition*

$$\mathcal{H}(\omega, \bar{x}(\omega), u, p(\omega)) - \mathcal{H}(\omega, \bar{x}(\omega), \bar{u}(\omega), p(\omega)) \geq 0 \text{ for all } u \in U, \text{ and a.e. } \omega \in \Omega. \tag{3.9}$$

Proof. By the second property in (3.8) we have

$$\begin{aligned} 0 \leq J(v) - J(\bar{u}) &= \int_\Omega \big(\ell(\cdot, x(v)) - \ell(\cdot, x(\bar{u})) + h(v) - h(\bar{u})\big)\, d\omega \\ &= \int_\Omega \big(\ell_x(\cdot, \bar{x})(x - \bar{x}) + h(v) - h(\bar{u})\big)\, d\omega + O(|x - \bar{x}|^2), \end{aligned}$$

where v is defined in (3.7) and $x = x(v)$. Utilizing the adjoint equation (3.6) we find that

$$\begin{aligned} &0 \leq J(v) - J(\bar{u}) \\ &\quad = -\langle (E + f_x(\cdot, \bar{x}, \bar{u}))(x - \bar{x}), p)\rangle_{X^*,X} + \int_\Omega \big(h(v) - h(\bar{u})\big)\, d\omega + O(|x - \bar{x}|^2). \end{aligned} \tag{3.10}$$

By the third property in (3.8) we have

$$\begin{aligned} 0 &= \langle E(x) + f(\cdot, x, v) - E(\bar{x}) - f(\cdot, \bar{x}, \bar{u}), p\rangle \\ &= \langle E(x - \bar{x}) + f_x(\cdot, \bar{x}, v)(x - \bar{x}) + f(\cdot, \bar{x}, v) - f(\cdot, \bar{x}, \bar{u}), p\rangle + O(|x - \bar{x}|^2) \\ &= \langle E(x - \bar{x}) + f_x(\cdot, \bar{x}, \bar{u})(x - \bar{x}) + f(\cdot, \bar{x}, v) - f(\cdot, \bar{x}, \bar{u}), p\rangle \\ &+ \langle f_x(\cdot, \bar{x}, v)(x - \bar{x}) - f_x(\cdot, \bar{x}, \bar{u})(x - \bar{x}), p\rangle + O(|x - \bar{x}|^2). \end{aligned}$$

By (3.4), (3.10) and the fourth property in (3.8) we find

$$
\begin{aligned}
0 &\le J(u) - J(\bar{u}) \\
&= \int_\Omega (f(\cdot,\bar{x},v) - f(\cdot,\bar{x},\bar{u}))p\,d\omega + \int_\Omega (h(v) - h(\bar{u}))\,d\omega \\
&\quad + \int_\Omega (f_x(\cdot,\bar{x},v) - f_x(\cdot,\bar{x},\bar{u}))(x-\bar{x})\,p\,d\omega + O(|x-\bar{x}|^2) \\
&\le \int_\Omega (f(\cdot,\bar{x},v) - f(\cdot,\bar{x},\bar{u}))p\,d\omega + \int_\Omega (h(v) - h(\bar{u}))\,d\omega \\
&\quad + (\int_\Omega |(f_x(\cdot,\bar{x},v) - f_x(\cdot,\bar{x},\bar{u}))\,p|^2\,d\omega)^{\frac{1}{2}}\,|x-\bar{x}|_{L^2} + O(|x-\bar{x}|^2).
\end{aligned}
$$

Now we restrict s to be a Lebesgue point of the mapping

$$\omega \to \big(h(\bar{u})(\omega),\ f(\omega,\bar{x}(\omega),\bar{u}(\omega))p(\omega)\big).$$

Let S denote the set of these Lebesgue points and note that $\text{meas}(S) = \text{meas}(\Omega)$. Dividing the last inequality by $|B(s,\delta)| > 0$, letting $\delta \to 0$, and using the first property in (3.8) we obtain

$$\mathcal{H}(s,\bar{x}(s),p(s),u) - \mathcal{H}(s,\bar{x}(s),p(s),\bar{u}) \ge 0 \text{ for all } u \in U$$

at Lebesgue points $s \in S$, and the claim follows.

□

3.2 Existence of a minimizer

After having derived a necessary optimality condition in the previous section, let us turn to the discussion of existence of solutions to (3.1)-(3.3) in the case that h involves the N_0 functional.

We start with some preliminary considerations. Let $h : \mathbb{R} \to \mathbb{R}$ be given by

$$h(u) = \frac{\alpha}{2}|u|^2 + \beta|u|_0, \tag{3.11}$$

where

$$|u|_0 = \begin{cases} 0 \text{ if } u = 0 \\ 1 \text{ if } u \ne 0. \end{cases}$$

The maximum principle established in Theorem 3.1 suggests to consider $\underset{u\in\mathbb{R}}{\operatorname{argmin}}\,(h(u)+qu)$. A short computation then shows that

$$\Phi(q) := \underset{u\in\mathbb{R}}{\operatorname{argmin}}\,(h(u)+qu) = \begin{cases} -\frac{q}{\alpha} & \text{for } |q| \geq \sqrt{2\alpha\beta} \\ 0 & \text{for } |q| < \sqrt{2\alpha\beta}. \end{cases} \tag{3.12}$$

Evaluating h at the minimum we obtain

$$h(\Phi(q)) + q\,\Phi(q) = \begin{cases} -\frac{1}{2\alpha}|q|^2 + \beta & \text{for } |q| \geq \sqrt{2\alpha\beta} \\ 0 & \text{for } |q| < \sqrt{2\alpha\beta}. \end{cases}$$

Clearly $-\Phi : \mathbb{R} \to \mathbb{R}$ is monotone, but it is not maximal monotone. For this reason we define

$$\tilde{\Phi}(q) \in \begin{cases} -\frac{q}{\alpha} & \text{for } |q| > \sqrt{2\alpha\beta} \\ 0 & \text{for } |q| < \sqrt{2\alpha\beta} \\ [\frac{-q}{\alpha}, 0] & \text{for } q = \sqrt{2\alpha\beta} \\ [0, -\frac{q}{\alpha}] & \text{for } q = -\sqrt{2\alpha\beta}. \end{cases}$$

The mapping $-\tilde{\Phi} : \mathbb{R} \to \mathbb{R}$ is maximal monotone. We also note that

$$h(\tilde{\Phi}(q)) + q\,\tilde{\Phi}(q) \in [0, \beta) \text{ for } q = |\sqrt{2\alpha\beta}|. \tag{3.13}$$

whereas $h(\Phi(q)) + q\,\Phi(q) = 0$ for $q = |\sqrt{2\alpha\beta}|$. Hence the effect of the extension of Φ to $\tilde{\Phi}$ on the u-part of the Hamiltonian along candidates of optimal solutions is bounded by β. This issue will be further addressed in Section 3.3.

In a natural way $\tilde{\Phi}$ defines an operator from $L^2(\Omega)$ to itself, which is also maximal monotone. It will be denoted by the same symbol.

Let us now consider the problem

$$\begin{cases} \min & \int_\Omega (\ell(\cdot, x) + \frac{\alpha}{2}|u|^2 + \beta|u|_0)\, d\omega, \\ \text{subject to} & Ex + f(x) + B(u) = 0, \quad u \in L^2(\Omega), \end{cases} \tag{3.14}$$

where $E \in \mathcal{L}(X, X^*)$, $B \in \mathcal{L}(L^2(\Omega), X^*)$ and ℓ, f satisfy (3.4) and (3.5). The case of control constraints will be considered separately in Section 4 below. We further define

$$F(u) = \int_\Omega \ell(\omega, x)\, d\omega$$

where $x = x(u) \in X$ for $u \in L^2(\Omega)$ is the unique solution to $Ex + f(x) + B(u) = 0$.

Theorem 3.2. *Suppose that there exists a solution $(x^*, p) \in X \times X$ to*

$$\begin{cases} Ex + f(x) + B\Phi(B^*p) = 0 \\ \\ (E + f'(x))^*p + \ell'(\cdot, x) = 0. \end{cases} \tag{3.15}$$

and set $u^* = \Phi(B^*p)$. If further

$$F(u) - F(u^*) \geq (B^*p, u - u^*) \quad \textit{for all } u \in L^2(\Omega), \tag{3.16}$$

then u^* is a solution to (3.14).

Proof. For $q = B^*p$ we have pointwise almost everywhere

$$h(u) - h(u^*) + q(u - u^*) = \begin{cases} \frac{\alpha}{2}|u - u^*|^2 & \text{for } |q| \geq \sqrt{2\alpha\beta} \\ \\ \frac{\alpha}{2}|u + \frac{q}{\alpha}|^2 + \beta - \frac{|q|^2}{2\alpha} & \text{for } |q| < \sqrt{2\alpha\beta}, u \neq 0 \\ \\ 0 & \text{for } |q| < \sqrt{2\alpha\beta}, u = 0. \end{cases}$$

This, together with (3.16) gives the desired result. □

We next turn to discuss conditions (3.15) and (3.16). For this purpose we consider the special case of a linear state equation

$$Ex + Bu = g \qquad (\text{i.e. } f(x, u) = Bu - g), \tag{3.17}$$

with $g \in X$ and assume that

$$x \to \ell(\omega, x) \text{ is convex.} \tag{3.18}$$

From the adjoint equation and

$$E(x(u) - x(u^*)) + B(u - u^*) = 0,$$

it follows that

$$F(u) - F(u^*) = (B^*p, u - u^*)_{L^2(\Omega)} + \int_\Omega (\ell(\cdot, x) - \ell(\cdot, x^*) - \ell'(x^*)(x - x^*))\, d\omega, \tag{3.19}$$

and hence the convexity assumption (3.18) for ℓ implies (3.16). If in addition to (3.17) we assume that $\ell_x = x - a$ with $a \in X$, then (3.15) reduces to

$$\begin{cases} Ex + B\Phi(B^*p) = g \\ \\ E^*p + x = a. \end{cases}$$

This saddle point problem may not have a solution, in general. We therefor introduce a relaxation by replacing the monotone operator $-\Phi$ by its maximal monotone extension $-\tilde{\Phi}$. This results in the system

$$\begin{cases} Ex + B\tilde{\Phi}(B^*p) \ni g \\ \\ E^*p + x = a. \end{cases} \tag{3.20}$$

Proposition 3.1. *If $E \in \mathcal{L}(X, X^*)$ is an isomorphism, then (3.20) admits a unique solution $(x, p) \in X \times X$ and there exists a constant K such that*

$$|(x, p)|_{X \times X} \le K|(g, a)|_{L^2 \times L^2} \ \ \textit{for every } (g, a) \in L^2(\Omega) \times L^2(\Omega).$$

Proof. Since E is an isomorphism it follows that $E^* \in \mathcal{L}(X, X^*)$ is an isomorphism as well. Let

$$D(E) = \{\varphi \in X : E\varphi \in L^2(\Omega)\}, \ \text{ and } D(E^*) = \{\varphi \in X : E^*\varphi \in L^2(\Omega)\}.$$

Endowed with the graph norm they are Hilbert spaces. Moreover $D(E)$ and $D(E^*)$ are dense in $L^2(\Omega)$ and $E \in \mathcal{L}(D(E), L^2(\Omega))$ and $E^* \in \mathcal{L}(D(E^*), L^2(\Omega))$ are isomorphism as well, see e.g. [11], page 19. Further define

$$D(EE^*) = \{\varphi \in X : E^*\varphi \in D(E), EE^*\varphi \in L^2(\Omega)\}.$$

EE^* is a closed monotone operator in $L^2(\Omega)$ satisfying

$$(EE^*\varphi, \varphi)_{L^2(\Omega)} \ge \kappa|\varphi|_X^2 \ \text{ for all } \varphi \in D(EE^*).$$

It follows that EE^* is maximal monotone in $L^2(\Omega)$. We set

$$\Psi(q) = \begin{cases} 0 & \text{for } |q| < \sqrt{2\alpha\beta} \\ \\ \frac{q^2}{2\alpha} & \text{for } |q| \ge \sqrt{2\alpha\beta}. \end{cases}$$

and note that Ψ is a proper convex function with $\partial(\Psi) = -\tilde{\Phi}$. Hence $-\tilde{\Phi}$ maximal monotone and the associated substitution operator from $L^2(\Omega)$ to $L^2(\Omega)$ is maximal monotone as well. Moreover $\Phi(B\cdot)$ is a proper monotone functional on $L^2(\Omega)$ and $\partial(\Psi(B^*\cdot)) = -B\tilde{\Phi}(B^*\cdot)$ is maximal monotone. Since the domain of $B\tilde{\Phi}B^*$ is all of $L^2(\Omega)$ it follows that $EE^* - B\tilde{\Phi}B^*$ is maximal monotone, see e.g. [3], Corollary 2.7.

For $a \in D(E)$ consider the equation

$$EE^*p - B\tilde{\Phi}B^*p \ni Ea - g. \tag{3.21}$$

Since

$$(EE^*\varphi - B\tilde{\Phi}B^*\varphi, \varphi)_{L^2} \geq \kappa|\varphi|_X^2 \to \infty \text{ for } |\varphi|_{L^2} \to \infty, \varphi \in D(EE^*),$$

it follows that $EE^* - B\tilde{\Phi}B^*$ is coercive and hence (3.21) admits a solution $p \in D(EE^*)$, see e.g. [3], page 31. It satisfies

$$|E^*p|_{L^2(\Omega)} \leq |a|_{L^2(\Omega)} + |g|_{L^2(\Omega)}.$$

With p determined, define x by $x = a - E^*p \in D(E)$. The pair (x,p) satisfies (3.20). Further there exists a constant K, independent of (a,g) such that

$$|Ex|_{L^2(\Omega)} \leq K(|a|_{L^2(\Omega)} + |g|_{L^2(\Omega)}), \quad \text{and } |(x,p)|_{X\times X} \leq K(|a|_{L^2(\Omega)} + |g|_{L^2(\Omega)}). \tag{3.22}$$

.

Now consider the case $a \in L^2(\Omega)$. Since $D(E)$ is dense in $L^2(\Omega)$, there exists a sequence $\{a_n\}$ in $D(E)$ such that $a_n \to a$ in $L^2(\Omega)$. Let (x_n, p_n) satisfy

$$\begin{cases} Ex_n + y_n = g, \quad \text{with } y_n \in B\tilde{\Phi}B^*p_n, \\ E^*p_n + x_n = a_n. \end{cases} \tag{3.23}$$

By (3.22) there exists a subsequence, denoted by the same expression such that $(x_n, p_n, y_n) \to (x, p, y)$ weakly in $X \times X \times L^2(\Omega)$, and $(x_n, p_n) \to (x, p)$ strongly in $L^2(\Omega) \times L^2(\Omega)$. Hence $\lim_{n\to\infty}(y_n, p_n)_{L^2}$ exists and by closed-ness property of the graph of the operator $-B\tilde{\Phi}B^*$ in $L^2(\Omega) \times L^2(\Omega)$ endowed with the strong-weak topology, [3] page 27, it follows that $y \in B\tilde{\Phi}B^*p$ and we can pass to the limit in (3.23) to obtain that (x,p) satisfies (3.20) and also (3.22).

To guarantee uniqueness of the solution, let (x,p) and $(\bar{x},\bar{p})$ denote two possibly different solutions and set $\delta x = x-\bar{x}$ and $\delta p = p-\bar{p}$. Then we have

$$E\,\delta x + B\tilde{\Phi}(B^*p) - B\tilde{\Phi}(B^*\bar{p}) \ni 0$$

$$E^*\,\delta p + \delta x = a.$$

Taking the inner product with $(\delta p, -\delta x)$ in the above equations and adding them up we obtain

$$[-|\delta x|^2_{L^2(\Omega)} + \left(B\tilde{\Phi}(B^*p) - B\tilde{\Phi}(B^*\bar{p}), \delta p\right)_{L^2(\Omega)}, 0] \ni 0,$$

which implies that $\delta x = \delta p = 0$. □

We summarize the above developments for the problem

$$\begin{cases} \min J(x,u) = \frac{1}{2}|x-a|^2_{L^2(\Omega)} + \frac{\alpha}{2}|u|^2_{L^2(\Omega)} + \beta N_0(u), \\ Ex + Bu = g, \quad u \in L^2(\Omega), \end{cases} \tag{3.24}$$

where $a \in L^2(\Omega)$, $\alpha > 0, g \in X^*$. While we consider here the case that the observation takes place on the whole domain, this is not essential for the results which will be presented. The tracking type functional could equally well be replaced by $\frac{1}{2}|x-a|^2_{L^2(\Omega_o)}$ with $\Omega_o \subsetneqq \Omega$.

Theorem 3.3. *Consider problem (3.24) with $E \in \mathcal{L}(X, X^*)$ an isomorphism. Let (x^*, p) be the unique solution to (3.20) and set $u^* = \tilde{\Phi}(B^*p)$. If meas $\{\omega : |B^*p(\omega)| = \sqrt{2\alpha\beta}\} = 0$, then (u^*, x^*) is a solution to (3.24) which satisfies the optimality system*

$$\begin{cases} Ex + Bu = g \\ E^*p + x - a = 0 \\ u^* = \begin{cases} \frac{-B^*p}{\alpha} & \text{for } |B^*p| \geq \sqrt{2\alpha\beta} \\ 0 & \text{for } |B^*p| < \sqrt{2\alpha\beta}, \end{cases} \end{cases} \tag{3.25}$$

with the last equality holding pointwise a.e.

Remark 3.2. If meas$\{\omega : |p(\omega)| = \sqrt{2\alpha\beta}\} > 0$, then (3.24) may not admit a solution, but we refer to $u^* \in \tilde{\Phi}(B^*p)$ as relaxed control. The effect of the N_0 term on this relaxed control as a function of β can be seen from (3.13) and will be considered further in Theorem 3.5 below.– The fact that the optimal control is identically equal to zero for p sufficiently small, justifies to call N_0 a sparsity enhancing functional.

Remark 3.3. Introducing $\lambda = \alpha u + B^*p$ the complementarity system in (3.25) can equivalently be expressed as

$$\begin{cases} \lambda = 0 & \text{if } |\lambda - \alpha u| > \sqrt{2\alpha\beta} \\ u = 0 & \text{if } |\lambda - \alpha u| \le \sqrt{2\alpha\beta}. \end{cases} \tag{3.26}$$

This form should be compared to the optimality system that was obtained in [14] for the discrete ℓ^0 problem $\min \frac{1}{2}(|Ax - a|_2^2 + \alpha|x|_2^2) + \beta|x|_0$, where $A \in \mathcal{L}(\ell^2)$, $|\cdot|_2$ denotes the norm in ℓ^2, and $|x|_0$ stands for the number of nonzero elements of $x \in \ell^2$. Here we add the $\alpha|x|_2^2$ part to the cost in the discrete formulation to match the continuous problem. It is not required for the analysis. Then, setting $\lambda_i = (A_i, A_i x^* - a) + \alpha x_i^*$, the optimality system in case of strict complementarity is given by

$$\begin{cases} \lambda_i = 0 & \text{if } \left|\lambda_i - (|A_i|^2 + \alpha)x_i^*\right| > \sqrt{2\beta}(|A_i|_2 + \sqrt{\alpha}) \\ \bar{x}_i = 0 & \text{if } \left|\lambda_i - (|A_i|^2 + \alpha)x_i^*\right| \le \sqrt{2\beta}(|A_i|_2 + \sqrt{\alpha}), \end{cases} \tag{3.27}$$

where $A_i = Ae_i$, with e_i the element in ℓ^2 which has 1 in the $i{-}th$ element and is 0 otherwise. Comparing to (3.26) we note that in the discrete formulation the tracking part of the cost sustains in the optimality condition whereas in (3.26) it does not. Thus for this class of problems, the order of discretization and optimization makes a significant difference.

3.3 Sufficiency

Theorem 3.4. *If $E \in \mathcal{L}(X, X^*)$ is an isomorphism and u^* is a solution to (3.24), then it is unique.*

Proof. Let u^* be an optimal solution with associated state $x^* = x(u^*)$ and let u be another control with state $x = x(u)$. Then we have

$$\begin{aligned} J(x,u) - J(x^*,u^*) &= \frac{1}{2}|x-a|^2_{L^2(\Omega)} - \frac{1}{2}|x^*-a|^2_{L^2(\Omega)} \\ &\quad + \frac{\alpha}{2}|u|^2_{L^2(\Omega)} - \frac{\alpha}{2}|u^*|^2_{L^2(\Omega)} + \beta(N_0(u) - N_0(u^*)) \\ &= (x^*-a, x-x^*)_{L^2} + \frac{1}{2}|x-x^*|^2_{L^2(\Omega)} + \alpha(u^*, u-u^*)_{L^2(\Omega)} \\ &\quad + \frac{\alpha}{2}|u-u^*|^2_{L^2(\Omega)} + \beta(N_0(u) - N_0(u^*))) \\ &= (B^*p + \alpha u^*, u-u^*)_{L^2(\Omega)} + \frac{\alpha}{2}|u-u^*|^2_{L^2(\Omega)} + \\ &\quad \frac{1}{2}|x-x^*|^2_{L^2(\Omega)} + \beta(N_0(u) - N_0(u^*)). \end{aligned}$$

Let us set $q = B^*p$ and define the sets

$$S^0 = \{x : |q| < \sqrt{2\alpha\beta}\} \text{ and } S^+ = \{x : |q| \geq \sqrt{2\alpha\beta}\}.$$

With respect to these sets we have, using that $\alpha u^* + q = 0$ on S^+,

$$\begin{aligned} J(x,u) - J(x^*,u^*) &= \beta \textstyle\int_{S^+}(\chi_{|u|\neq 0} - \chi_{|u^*|\neq 0})d\omega \\ &+ \textstyle\int_{S^0}(qu + \beta\chi_{u\neq 0})d\omega + \frac{\alpha}{2}|u-u^*|^2_{L^2(\Omega)} + \frac{1}{2}|x-x^*|^2_{L^2(\Omega)}. \end{aligned} \tag{3.28}$$

We further set

$$S_0^+ = \{x \in S^+ : u = 0\}, \text{ and } S_0^0 = \{x \in S^0 : u \neq 0\}.$$

Then we have using $|q| \geq \sqrt{2\alpha\beta}$ on S^+

$$\begin{aligned} &J(x,u) - J(x^*,u^*) \geq \frac{\alpha}{2}\int_{S^+\setminus S_0^+} |u-u^*|^2 d\omega \\ &+ \int_{S_0^+} (\frac{1}{2\alpha}|q|^2 - \beta)\, d\omega + \int_{S_0^0} (qu + \beta + \frac{\alpha}{2}|u|^2)d\omega + \frac{1}{2}|x-x^*|^2_{L^2(\Omega)} \\ &\geq \frac{\alpha}{2}\int_{S^+\setminus S_0^+} |u-u^*|^2 d\omega + \int_{S_0^0} (\frac{1}{2}|\sqrt{\alpha}u + \frac{q}{\sqrt{\alpha}}|^2 + \beta - \frac{1}{2}\frac{|q|^2}{\alpha})d\omega + \frac{1}{2}|x-x^*|^2_{L^2(\Omega)}. \end{aligned} \tag{3.29}$$

If meas$(S_0^0) \neq 0$, then $J(x,u) - J(x^*,u^*) > 0$, since $|q| < \sqrt{2\alpha\beta}$ a.e. on S_0^0. Otherwise $u = u^*$ a.e. on S^0 and

$$J(x,u) - J(x^*,u^*) \geq \frac{\alpha}{2}\int_{S^+\setminus S_0^+} |u-u^*|^2\, d\omega + \frac{1}{2}|x-x^*|^2_{L^2(\Omega)sti}.$$

If $\mathrm{meas}(S_0^+) = 0$ and $J(x,u) = J(x^*,u^*)$ then $u = u^*$ a.e. on S^+ and hence $u = u^*$ a.e. in Ω. Otherwise $\mathrm{meas}(S_0^+) > 0$ and $u = 0$ on S_0^+, $u^* \neq 0$ on S_0^+. Consequently $x \neq x^*$ and again $J(x,u) > J(x^*,u^*)$. This implies that u^* is the unique global minimum of (3.24).

□

In the following result we quantify the quality of the relaxed optimal control that is obtained from the optimality system (3.20) as suboptimal solution to (3.24).

Theorem 3.5. *Under the assumptions of Theorem 3.3, if* $u^* \in \tilde{\Phi}(B^*p)$ *with* (x^*, p) *the solution to* (3.20)*, then*

$$J(x(u),u) > J(x(u^*),u^*) - \beta\, meas\, \mathcal{S}, \tag{3.30}$$

for every $u \in L^2(\Omega)$ *with* $u \neq u^*$*, where* $\mathcal{S} = \{\omega : |B^*p| = \sqrt{2\alpha\beta}\}$.

Proof. From the adjoint equation we have

$$\begin{aligned}
&J(x(u),u) - J(x(u^*),u^*)\\
&= (q, u-u^*)_{L^2(\Omega)} + \textstyle\int_\Omega (h(u)-h(u^*))\,d\omega + \frac{1}{2}|x-x^*|^2_{L^2(\Omega)}\\
&= (q, u-u^*)_{L^2(\mathcal{S})} + \textstyle\int_{\mathcal{S}} (h(u)-h(u^*))\,d\omega + (q, u-u^*)_{L^2(\Omega\setminus\mathcal{S})}\\
&+ \textstyle\int_{\Omega\setminus\mathcal{S}} (h(u)-h(u^*))\,d\omega + \frac{1}{2}|x-x^*|^2_{L^2(\Omega)}\\
&\geq (q,u)_{L^2(\mathcal{S})} + \textstyle\int_{\mathcal{S}} h(u)\,d\omega - (q,u^*)_{L^2(\mathcal{S})} - \int_{\mathcal{S}} h(u^*)\,d\omega,
\end{aligned}$$

where for the last estimate we can proceed as in the proof of Theorem 3.4. A simple computation and (3.13) imply that

$$J(x(u),u) - J(x(u^*),u^*) \geq -(\sqrt{2\alpha\beta}, u^*)_{L^2(\mathcal{S})} - \textstyle\int_{\mathcal{S}} h(u^*)\,d\omega > -\beta\,\mathrm{meas}\,\mathcal{S},$$

and the inequality is strict, if $\mathrm{meas}\,\mathcal{S} > 0$. This implies (3.30) with $>$ replaced by $\geq$. Strict inequality holds since it was already obtained in Theorem 3.4 if $\mathrm{meas}\,\mathcal{S} = 0$ and if $\mathrm{meas}\,\mathcal{S} > 0$ then it follows from (3.13). □

3.4 Example

Consider the optimal control problem;

$$\begin{aligned} &\min \quad \int_0^1 (\tfrac{1}{2}|x|^2 + \tfrac{\alpha}{2}|u|^2 + |u|_0)\,dt \\ &\tfrac{dx}{dt} = u, \;\; x(0) = x_0 > 0, \quad u \in L^2(0,1). \end{aligned} \tag{3.31}$$

Setting $y = x + x_0$ this problem can equivalently be formulated as

$$\min \quad \int_0^1 (\frac{1}{2}|y + x_0|^2 + \frac{\alpha}{2}|u|^2 + |u|_0)\,dt$$

subject to

$$\frac{dy}{dt} = u, \;\; y(0) = 0.$$

It fits into the general setup of the previous sections with $\Omega = (0,1)$, $X = \{y \in W^{1,2}(\Omega) : y(0) = 0\}$, $E = \frac{d}{dt}$ and $B = -I$. The operator E is not an isomorphism, but this is of no concern, since we obtain existence of the optimality system directly.

Returning to (3.31), the optimality system is given by

$$\begin{cases} \frac{d}{dt}x + \tilde{\Phi}(p) \ni 0, \; x(0) = x_0 \\ \frac{d}{dt}p = x, \; p(1) = 0, \end{cases} \tag{3.32}$$

and for the optimal control we have $u \in \tilde{\Phi}(p)$. Since $x_0 > 0$ there exists a maximal $\hat{t} \in (0,1]$ such that $x(t) > 0$ for $t \in \mathcal{J} := [0, \hat{t})$, and $x(\hat{t}) = 0$ with $\hat{t} \in (0,1)$ or $\hat{t} = 1$. Assume for the moment that a solution to the optimality system exists and the $\hat{t} \in (0,1)$. Then necessarily $x = u = p$ on $\mathcal{J}$. By continuity of p there exists a neighborhood of $\hat{t}$ within which $u = 0$ and hence $\frac{d}{dt}x$ =constant $= 0$. This contradicts the maximality of $\hat{t}$. Hence $x > 0$, $\frac{d}{dt}p > 0$ and $p < 0$ on $[0,1)$. In particular, if a solution to (3.32) exists, then there exists a solution with $\tilde{\Phi}$ replaced by Φ and the structure of the optimality system is given by

$$\begin{cases} \dfrac{dx}{dt} = \begin{cases} \frac{p}{\alpha} \text{ on } [0, t^*) \\ 0 \text{ on } (t^*, 1], \end{cases} \qquad x(0) = x_0, \\ \dfrac{dp}{dt} = x, \quad p(t^*) = -\sqrt{2\alpha} \text{ and } p(1) = 0, \end{cases}$$

and the optimal solution is given by

$$u^* = \frac{p}{\alpha} \text{ on } [0, t^*) \quad \text{and} \quad 0 \text{ on } (t^*, 1].$$

Here we assume that $t^* \in (0, 1)$. This implies that for constants A, B we have

$$x(t) = Ae^t + Be^{-t} \text{ on } [0, t^*] \quad \text{and} \quad x(t) = x^* = Ae^{t^*} + Be^{-t^*} \text{ on } [t^*, 1].$$

$$p(t) = Ae^t - Be^{-t} \text{ on } [0, t^*] \quad \text{and} \quad p(t) = x^*(t-1) \text{ on } [t^*, 1]$$

where $A + B = x_0$. Using continuity of p this reduces to

$$-Ae^{t^*} + (x_0 - A)e^{-t^*} = \sqrt{2\alpha} = -p(t^*),$$

$$(1 - t^*)(Ae^{t^*} + (x_0 - A)e^{-t^*}) = \sqrt{2\alpha},$$

and thus

$$A = \frac{x_0 e^{-t^*} - \sqrt{2\alpha}}{e^{t^*} + e^{-t^*}},$$

$$(1 - t^*)(x_0 e^{-t^*} + \tanh(t^*)(x_0 e^{-t^*} - \sqrt{2\alpha})) = \sqrt{2\alpha}.$$

For $x_0 = 4$ and $\alpha = 1$ a numerical calculation gives $t^* \sim 0.5077$. Thus we have found a solution to the optimality system, without relaxation. The associated control is optimal, it has one jump, and it is unique.

3.5 Primal-Dual active set methods

The complementary condition (3.26) suggests a primal-dual active set strategy to solve (3.24). It does not realize the set-valued nature of (3.32), but it converges within a few iterations for moderate values of $\frac{\beta}{\alpha}$.

Primal-Dual Active Set Method

- Initialize p^0 and set $n = 0$.
- Solve for $(x^{n+1}, u^{n+1}, p^{n+1})$

  $$Ex^{n+1} + Bu^{n+1} = g, \quad E^* p^{n+1} + \ell'(x^{n+1}) = 0,$$

  and

  $$u^{n+1} = -\tfrac{1}{\alpha} B^* p^{n+1} \quad \text{on } \{|B^* p^n| \geq \sqrt{2\alpha\beta}\},$$

  $$u^{n+1} = 0 \qquad \text{on } \{|B^* p^n| < \sqrt{2\alpha\beta}\}.$$

Table 1:

β	.5	.1	.05	.01	.005	.001
no. of iterates	1	5	4	4	2	2
N_0	0	1135	2852	7296	8853	12090
$\lvert\mathcal{C}_1\rvert$	0	11	15	47	57	89

- The stopping criterion satisfied and Stop, or
  Set $n = n + 1$, and return to the second step.

The stopping criterion we chose utilizes the critical set

$$\mathcal{C}_1 = \{\sqrt{2\alpha\beta} - h^2 \le |B^* p^n| \le \sqrt{2\alpha\beta} + h^2\}$$

where h denotes the mesh-size of the discretization of the continuous operators, and the algorithm is stopped as soon as the discretized versions of

$$|Ex^n + Bu^n - g|_{L^\infty(\Omega\setminus\mathcal{C}_1)} \le \ \mathrm{tol}_1 \quad \text{and} \quad |B^* p^n|_{L^\infty(\mathcal{C}_1)} \le (1 + \mathrm{tol}_2)\sqrt{\tfrac{2\beta}{\alpha}}$$

for a given tolerances tol_i are satisfied.

We briefly report on a numerical example with $E = -\Delta$, with Dirichlet boundary conditions, $B = I$, $g = 0$, and $X = H^1_0(\Omega)$, Ω the unit square and discretization based on finite differences with respect to a uniform mesh $h = 1/N$ to solve (3.24).

Example 3.1 We choose $a = 10\,\omega_1 \sin(5\omega_1) \cos(7\omega_2)$, and give results for $N = 128$, $\alpha = .01$ and a sequence of β values, with $\mathrm{tol}_1 = 10^{-11}$, $\mathrm{tol}_2 = \frac{1}{N}$.

Here and below the algorithms are always initialized by solving the optimal control problem with $\beta = 0$.

In Table 1, N_0 denotes the number of interior nodes which are different from zero and $|\mathcal{C}_1|$ stands for the number of nodes in $\mathcal{C}_1$. For $N = 128$ the number of interior nodes is 16129. For $\beta = .5$ we obtain $N_0 = 0$, i.e. we have maximal sparsity.

It is consistent with our expectation that N_0 increases as β decreases. If the iteration is continued after the stopping criterion is reached then the iterates stay constant except for the case $\beta = .1$. If the iterates stay constant then an exact solution of the discretized problem is found. In case $\beta = .1$

the algorithm is periodic with two states, each of which satisfy the stopping criterion.

The situation that the algorithm enters into a periodic behavior also typically arises for cases when the fraction $\frac{\alpha}{\beta}$ is smaller than those used in Table 1. This comes as no surprise, because in the Primal-Dual Active Set Method proposed above we have not yet accounted for the fact that the graph Φ must be extended to be maximal monotone to guarantee existence.

To compute approximate solutions to (3.24) if $\frac{\beta}{\alpha}$ is large we utilize a regularized form of the operator Φ which appears in the optimality condition. It is given by:

$$\Phi^\epsilon(q) = \begin{cases} \frac{-q}{\alpha} & \text{for } |q| > \sqrt{2\alpha\beta} + \epsilon \\ 0 & \text{for } |q| < \sqrt{2\alpha\beta} - \epsilon \\ -\frac{(\sqrt{2\alpha\beta}+\epsilon)q}{2\alpha\epsilon} + \operatorname{sign}(q)\frac{2\alpha\beta-\epsilon^2}{2\alpha\epsilon} & \text{for } \sqrt{2\alpha\beta} - \epsilon \le |q| \le \sqrt{2\alpha\beta} + \epsilon. \end{cases}$$

We observe that $-\Phi^\epsilon$ is maximal monotone. Accordingly the active set strategy is modified and we arrive at the following algorithm.

Regularized Primal-Dual Active Set Method

Here the second bullet of the Primal-Dual Active Set Method is replaced by:

- Solve for $(x^{n+1}, u^{n+1}, p^{n+1})$

$$Ex^{n+1} + Bu^{n+1} = g, \quad E^*p^{n+1} + \ell'(x^{n+1}) = 0,$$

and

$$u^{n+1} = \begin{cases} -\frac{1}{\alpha}B^*p^{n+1} & \text{on } \{|B^*p^n| \ge \sqrt{2\alpha\beta} + \epsilon\}, \\ 0 & \text{on } \{|B^*p^n| < \sqrt{2\alpha\beta} - \epsilon\}, \\ -\frac{(\sqrt{2\alpha\beta}+\epsilon)B^*p^{n+1}}{2\alpha\epsilon} + \operatorname{sign}(B^*p^n)\frac{2\alpha\beta-\epsilon^2}{2\alpha\epsilon} & \\ \qquad \text{on } \{\sqrt{2\alpha\beta} - \epsilon \le |B^*p^n| \le \sqrt{2\alpha\beta} + \epsilon\}. \end{cases}$$

Table 2:

β	.5	.1	.05	.01	.005	.001
no. of iterates	2	5	4	5	3	3
N_0	0	1134	2852	7295	8852	12089
$\lvert\mathcal{C}_2\rvert$	0	1	0	1	3	1

Table 3:

β	5	1	.5	.1	.01
no. of iterates	32	10	6	6	4
N_0	1115	4505	6370	10098	13404
$\lvert\mathcal{C}_2\rvert$	3	20	16	21	33

Example 3.2 This and the following example are computed with the regularized algorithm. The algorithm is stopped as soon as two consecutive iterates coincide and the exact discretized solution is obtained.

First we consider exactly the same specifications as in Example 3.1. The results that are obtained with the regularized algorithm with $\epsilon = 10^{-6}$ are depicted in Table 2. We note that N_0 is very similar to the results obtained with the unregularized algorithm in spite of the fact that the procedure for obtaining critical set, where $|p| \sim \sqrt{2\alpha\beta}$, is different. In Table 2, $|\mathcal{C}_2|$ stands for the number of nodes in $\mathcal{C}_2 = \{\sqrt{2\alpha\beta} - \epsilon \leq |B^* p^n| \leq \sqrt{2\alpha\beta} + \epsilon\}$.

Subsequently we made tests with $\alpha = 10^{-4}$, and $\epsilon = 10^{-5}$. The results for a series of β-values are given in Table 3. Concerning the dependence of the solution on ϵ we tested with $\alpha = 10^{-3}$ and found that for ϵ in the range 10^{-5} to 10^{-3} the number of zero nodes ranges between 15148 and 15013. We also confirmed that in the final iteration, the control satisfies $|u| \leq \frac{\sqrt{2\alpha\beta}+\epsilon}{\alpha}$ over the region where $\{\sqrt{2\alpha\beta} - \epsilon \leq |p| \leq \sqrt{2\alpha\beta} + \epsilon\}$ which accounts for closing the graph of Φ at $\sqrt{2\alpha\beta}$.

Example 3.3 In this example the control and observation occupy only part of the domain and they are non overlapping. We choose $\Omega_c = (0,1) \times (0, \frac{1}{3})$ and the observation is restricted to $\Omega_o = (0,1) \times (\frac{3}{4}, 1)$. The choice for a and $n = 128$ are as in the previous examples and $\alpha = 10^{-4}$, $\epsilon = 10^{-6}$. The number of nodes that lie in the control domain is 5534. The numbers for N_0 and $|\mathcal{C}_2|$ in Table 4 refer only to this set.

Table 4:

β	.05	.01	.001	.0001
no. of iterates	2	2	3	3
N_0	0	535	2267	3780
$\lvert\mathcal{C}_2\rvert$	0	2	9	17

It should be noted that for the examples presented here, the corresponding optimal controls are zero on sets which contain interior points. Thus these type of sparse controls differ from those obtained by with $L^1(\Omega)$ (or more precise, measure valued cost-functionals). The latter are more rough, and in the case that the desired states contain e.g. objects with edges they are of co-dimension one type. This is not the case for controls computed with $L^0(\Omega)$ combined with $L^2(\Omega)$ cost functionals.

4 L^0: optimal control with control constraints

We return to problem (3.24) with constraints on the controls:

$$\begin{cases} \min & \int_\Omega (\ell(\cdot,x) + \beta|u|_0)\, d\omega, \\ \text{subject to} & Ex + f(x) + B(u) = 0, \\ & u \in U_{ad} = \{u \in L^2(\Omega) : a \le u(\omega) \le b, a.e.\}, \end{cases} \tag{4.1}$$

where $a < 0 < b$.

We shall demonstrate that, except possibly at switching points, the optimal control can only achieve that a, b or 0.

As in Section 3.2 we start with preliminaries involving the Hamiltonian. We define $h : \mathbb{R} \to \mathbb{R}$ by

$$h(u) = \beta|u|_0 + \chi_{[a,b]}, \tag{4.2}$$

where $\chi_{[a,b]}$ denotes the indicator function of the closed interval $[a,b]$. We find

$$\Phi(q) := \underset{u\in\mathbb{R}}{argmin}\,(h(u)+qu) = \begin{cases} 0 & \text{for } -\frac{\beta}{b} < q < -\frac{\beta}{a} \\ b & \text{for } q \leq -\frac{\beta}{b} \\ a & \text{for } q \geq -\frac{\beta}{a}, \end{cases} \tag{4.3}$$

and

$$h(\Phi(q)) + q\,\Phi(q)) = \begin{cases} 0 & \text{for } -\frac{\beta}{b} < q < -\frac{\beta}{a} \\ \beta + qb & \text{for } q \leq -\frac{\beta}{b} \\ \beta + qa & \text{for } q \geq -\frac{\beta}{a}. \end{cases}$$

Again $-\Phi : \mathbb{R} \to \mathbb{R}$ is monotone, but not maximal monotone. The maximal monotone extension is given by

$$\tilde{\Phi}(q) \in \begin{cases} 0 & \text{for } -\frac{\beta}{b} < q < -\frac{\beta}{a} \\ b & \text{for } q < -\frac{\beta}{b} \\ a & \text{for } q > -\frac{\beta}{a}, \\ [0,b] & \text{for } q = -\frac{\beta}{b} \\ [a,0] & \text{for } q = -\frac{\beta}{a}. \end{cases} \tag{4.4}$$

We also have

$$h(\tilde{\Phi}(q)) + q\,\Phi(q) \in [0,\beta) \text{ for } q \in \{-\frac{\beta}{b}, -\frac{\beta}{a}\}. \tag{4.5}$$

In passing let us mention that for the case that $U = [0,b]$ the resulting feedback operator has the form

$$\tilde{\tilde{\Phi}}\,(q) \in \begin{cases} 0 & \text{for } -\frac{\beta}{b} < q \\ b & \text{for } q < -\frac{\beta}{b} \\ [0,b] & \text{for } q = -\frac{\beta}{b}. \end{cases}$$

Thus, according to the maximum principle (3.9), the optimal controls, aside from the switching points $\{\frac{-\beta}{b}, \frac{-\beta}{a}\}$ can only assume three, respectively, two states.

Theorem 4.1. *Consider problem* (4.1) *with* $E \in \mathcal{L}(X, X^*)$ *an isomorphism. Let* (x^*, p) *be the unique solution to* (3.20)*, with* $\tilde{\Phi}$ *given in* (4.4) *and set* $u^* = \tilde{\Phi}(B^*p)$*. If* $meas\,\mathcal{S} = 0$ *with* $\mathcal{S} = \{\omega : B^*p(\omega) \in \{-\frac{\beta}{b}, -\frac{\beta}{a}\}\}$*, then* (4.1) *admits a solution* (u^*, x^*) *which satisfies the optimality system*

$$\begin{cases} Ex + Bu = g \\ E^*p + x - a = 0 \\ u^* = \begin{cases} 0 & \text{for } -\frac{\beta}{b} < B^*p < -\frac{\beta}{a} \\ b & \text{for } B^*p \le -\frac{\beta}{b} \\ a & \text{for } B^*p \ge -\frac{\beta}{a}, \end{cases} \end{cases}$$

with the last equality holding pointwise a.e. - If $meas\,\mathcal{S} \neq 0$*, then* u^* *is a relaxed control.*

Proof. The arguments of the proof of Proposition 3.1 guarantee the existence of a solution to (3.20) with $\tilde{\Phi}$ given in (4.4).

It suffices to argue that (x^*, u^*) is indeed a solution to (4.1). For this purpose we estimate

$$h(u) - h(u^*) + B^*p(u - u^*) \ge 0 \quad \text{for } u \in [a, b].$$

Using (3.16) and (3.19) the claim follows. □

Concerning the relaxed solution, we again have an estimate analogous to that in Theorem 3.5.

Theorem 4.2. *Under the assumptions of Theorem 4.1, if* $u^* \in \tilde{\Phi}(B^*p)$ *with* (x^*, p) *the solution to* (3.20)*, then*

$$J(x(u), u) > J(x(u^*), u^*) - \beta\, meas\,\mathcal{S}, \tag{4.6}$$

for every $u \in U_{ad}$ *with* $u \neq u^*$*, where* $\mathcal{S} = \{\omega : B^*p(\omega) \in \{-\frac{\beta}{b}, -\frac{\beta}{a}\}\}$*.*

Proof. Let $q = B^*p$ and $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 = \{q = -\frac{\beta}{b}\} \cup \{q = -\frac{\beta}{a}\}$. Following the proof of Theorem 3.5 we find for $u \in U_{ad}$

$$J(x(u), u) - J(x(u^*), u^*) = (q, u - u^*)_{L^2(\Omega\setminus\mathcal{S})} + \beta \int_{\Omega\setminus\mathcal{S}} (|u|_0 - |u^*|_0)\, d\omega +$$

$$(q, u - u^*)_{L^2(\mathcal{S})} + \beta \int_{\mathcal{S}} (|u|_0 - |u^*|_0)\, d\omega$$

$$\ge (-\tfrac{\beta}{b}, u)_{L^2(\mathcal{S}_1)} + \beta \int_{L^2(\mathcal{S}_1)} |u|_0\, d\omega + (-\tfrac{\beta}{a}, u)_{L^2(\mathcal{S}_2)} + \beta \int_{L^2(\mathcal{S}_2)} |u|_0\, d\omega$$

$$+(\tfrac{\beta}{b}, u^*)_{L^2(\mathcal{S}_1)} + (\tfrac{\beta}{a}, u^*)_{L^2(\mathcal{S}_2)} - \beta \int_{L^2(\mathcal{S})} |u^*|_0\, d\omega,$$

and the inequality is strict, unless meas $\mathcal{S}$ = meas Ω. The first four summands of the last expression combined are non-negative, the fifth and sixth summand as well are non-negative, and therefore we have

$$J(x(u),u) - J(x(u^*),u^*) \geq -\beta \operatorname{meas} \mathcal{S}.$$

For $u \neq u^*$ and $\operatorname{meas} \mathcal{S} < meas\, \Omega$ this inequality is strict as already noted above. For $\operatorname{meas} \mathcal{S} > 0$ we have $(\frac{\beta}{b}, u^*)_{L^2(\mathcal{S}_1)} + (\frac{\beta}{a}, u^*)_{L^2(\mathcal{S}_2)} - \beta \int_{L^2(\mathcal{S})} |u^*|_0 \, d\omega > -\beta \operatorname{meas} \mathcal{S}$ by (4.5). This concludes the proof. □

5 L^0: general case with regularization

In Section 3 we considered optimization problems involving N_0 with the specific structure of optimal control problems. Here we turn to a general class of problems. To guarantee existence we utilize an $H^1(\Omega)$ regularization term. We also utilize a smoothing of the $|\cdot|_0$ function given by

$$|x|_0^\epsilon = \begin{cases} 1 & |x| \geq \epsilon \\ \frac{|x|}{\epsilon} & |x| \leq \epsilon, \end{cases}$$

with $\varepsilon > 0$. For $\alpha > 0$, $\beta > 0$ we consider the problem

$$\min \quad j(u) + \frac{\alpha}{2} |\nabla u|^2_{L^2} + \beta \int_\Omega |u|_0^\varepsilon \, d\omega, \text{ over } u \in X = H^1(\Omega), \tag{5.1}$$

where $j \in C^1(X, \mathbb{R})$ is bounded from below, weakly lower semi-continuous and

$$u \to j(u) + \frac{\alpha}{2} |\nabla u|^2 \quad \text{is radially unbounded,} \tag{5.2}$$

i.e. $|u_n|_X \to \infty$ implies that $j(u_n) + \frac{\alpha}{2}|\nabla u_n|^2 \to \infty$.

Setting $j(u) = \frac{1}{2}|E^{-1}(g - Bu) - a|^2_{L^2}$ relates to problem (3.24), this time with a different regularisation term.

The sets of directional derivatives of $|\cdot|_0^\varepsilon$ are given by

$$(|x|_0^\epsilon)' = \begin{cases} 0 & |x| > \varepsilon \\ \frac{\operatorname{sgn}(x)}{\varepsilon} & 0 < |x| < \varepsilon \\ \{0, \frac{\operatorname{sgn}(x)}{\varepsilon}\} & x = \pm\, \varepsilon \\ \{\pm \frac{1}{\varepsilon}\} & x = 0. \end{cases}$$

Theorem 5.1. *Problem* (5.1) *admits a solution* u_ε. *It satisfies the optimality condition*

$$j'(u_\varepsilon) + \lambda_\varepsilon - \alpha\Delta u_\varepsilon = 0, \quad \lambda_\varepsilon \in \beta\,(|u|_0^\varepsilon)'.$$

Proof. Since j is bounded below there exists a minimizing sequence, which is bounded in X due to (5.2). We recall that $H^1(\Omega)$ is compactly embedded. Together with Lebesgue's bounded convergence theorem the existence proof follows with standard arguments. Computing directional derivatives, the necessary optimality condition follows. □

Theorem 5.2. *Every weak subsequential limit* $\bar{u} \in X$ *of* $\{u_\varepsilon\}$ *as* $\varepsilon \to 0^+$ *is a solution to*

$$\min \quad j(u) + \frac{\alpha}{2}\,|\nabla u|^2_{L^2(\Omega)} + \beta \int_\Omega |u|_0\,d\omega, \text{ over } u \in X = H^1(\Omega). \tag{5.3}$$

If moreover $u \to j'(u)$ *is continuous from the weak topology in* X *to the weak topology in* X^*, *then* $\lambda_\varepsilon \to \bar{\lambda}$ *weakly in* X^* *and*

$$\begin{cases} j'(\bar{u}) + \bar{\lambda} - \alpha\Delta\bar{u} = 0 \text{ in } X^*, \\ \lambda^\epsilon(\omega) \to \bar{\lambda} = 0 \text{ a.e. on } \{|\bar{u}| > 0\}, \\ \langle \bar{\lambda}, \phi\,\bar{u}\rangle_{X^*,X} \geq 0 \text{ for all } \phi \in C^1(\omega) \text{ with } \phi \geq 0. \end{cases} \tag{5.4}$$

Proof. Choose a subsequence of $\{u_\varepsilon\}_{\varepsilon>0}$ and $\bar{u} \in X$ such that $u_\varepsilon \to \bar{u}$ weakly in X, strongly in $L^2(\Omega)$ and pointwise almost everywhere. Then by Fatou's lemma

$$\begin{aligned} &\liminf_{\varepsilon\to 0^+}(j(u_\varepsilon) + \tfrac{\alpha}{2}|\nabla u^\varepsilon|^2_{L^2(\Omega)} + \beta \textstyle\int_\Omega |u^\varepsilon|_0^\varepsilon\,d\omega) \\ &\quad \geq j(\bar{u}) + \tfrac{\alpha}{2}|\nabla\bar{u}|^2_{L^2(\Omega)} + \liminf_{\varepsilon\to 0^+} \beta \textstyle\int_\Omega |u^\varepsilon|_0^\varepsilon\,d\omega \\ &\quad \geq \; j(\bar{u}) + \tfrac{\alpha}{2}|\nabla\bar{u}|^2_{L^2(\Omega)} + \beta \textstyle\int_\Omega |\bar{u}|_0\,d\omega. \end{aligned}$$

On the other hand for every $u \in X$ we have by Lebesgue's bounded convergence theorem

$$\begin{aligned} &\limsup_{\varepsilon\to 0^+}(j(u_\varepsilon) + \tfrac{\alpha}{2}|\nabla u^\varepsilon|^2_{L^2(\Omega)} + \beta \textstyle\int_\Omega |u^\varepsilon|_0^\varepsilon\,d\omega) \\ &\quad \leq j(u) + \tfrac{\alpha}{2}|\nabla u|^2_{L^2(\Omega)} + \limsup_{\varepsilon\to 0^+} \beta \textstyle\int_\Omega |u|_0^\varepsilon\,d\omega \\ &\quad \leq j(u) + \tfrac{\alpha}{2}|\nabla u|^2_{L^2(\Omega)} + \beta \textstyle\int_\Omega |u|_0\,d\omega, \end{aligned}$$

and thus $\bar{u}$ is a solution to (5.3).

By the regularity assumption on $u \to j'(u)$ we have that $\lambda_\varepsilon \to \bar{\lambda}$ weakly in $X*$ and the first equation in (5.4) follows. Since $u_\varepsilon \to \bar{u}$ we have the second equation in (5.4). Finally taking the limit in $\langle \bar{\lambda}_\varepsilon, \phi\, u_\varepsilon \rangle_{X^*,X} \geq 0$ we obtain the last statement of (5.4).

□

Remark 5.1. From Tchebycheff's inequality we have

$$\text{meas}\{\omega : |u_\epsilon(\omega) - \bar{u}(\omega)| > \delta\} \leq \tfrac{1}{\delta^2} |u_\epsilon - \bar{u}|^2_{L^2(\Omega)}$$

for every $\delta > 0$. In particular, this implies that

$$\text{meas}\{\omega : \bar{u}(\omega) = 0, |u_\epsilon(\omega)| > \delta\} \leq \tfrac{1}{\delta^2} |u_\epsilon - \bar{u}|^2_{L^2(\Omega)} \to 0.$$

6 L^p: Local regularisation and monotone scheme

In this section we turn to minimization problems involving N_p functionals with $p \in (0, 1)$. We restrict ourselves from the beginning to a class of problems for which we shall prove the convergence of a monotone scheme, and consider

$$\min\ \mathcal{J}(u) = j(u) + \frac{\alpha}{2} |\nabla u|^2_{L^2(\Omega)} + \beta\, N_p(u) \quad \text{over}\ \ u \in H^1(\Omega) \cap \mathcal{C}, \qquad (P_s)$$

where $j : H^1(\Omega) \to \mathbb{R}$ is weakly lower semi-continuous, $\alpha > 0, \beta > 0$, and $\mathcal{C} \subset L^2(\Omega)$ is closed and convex.

To cope with the singularity N_p, $p \in (0, 1)$ at the origin we consider a family of regularized problems. For this purpose we introduce the concave functions $\psi_\varepsilon : \mathbb{R}^+ \to \mathbb{R}^+$ by

$$\psi_\varepsilon(t) = \begin{cases} \frac{p}{2} \frac{t}{\varepsilon^{2-p}} + (1 - \frac{p}{2})\varepsilon^p & \text{for } t \in [0, \varepsilon^2] \\ t^{\frac{p}{2}} & \text{for } t \in (\varepsilon^2, \infty), \end{cases} \tag{6.1}$$

where $\varepsilon > 0$ and note that

$$\psi'_\varepsilon(t) = \frac{p}{2} \frac{1}{\max(\varepsilon^{2-p}, t^{\frac{2-p}{2}})}, \quad \text{for } t > 0,$$

in particular $\psi_\varepsilon \in C^1$. We note that while ψ_ε is concave this is not the case for $t \to \psi_\varepsilon(t^2)$

The family of regularized problems that we consider is then defined by

$$\min \mathcal{J}_\varepsilon(u) \ = j(u) + \frac{\alpha}{2}\,|\nabla u|^2_{L^2(\Omega)} + \beta\, N_{p,\varepsilon}(u), \text{ over } u \in H^1(\Omega) \cap \mathcal{C}, \quad (P_{s,\varepsilon})$$

where

$$N_{p,\varepsilon}(u) = \int_\Omega \psi_\varepsilon(|u|^2)\, dx.$$

We can check that for any $\varepsilon > 0$

$$N_p(u) \le N_{p,\varepsilon}(u) \le \max(N_p(u), \varepsilon^p|\Omega|). \tag{6.2}$$

The following lemma will be useful below.

Lemma 6.1. *If $u_n \to u$ in $L^1(\Omega)$, then*

$$N_p(u_n) \to N_p(u) \quad \textit{and} \quad N_{p,\varepsilon}(u_n) \to N_{p,\varepsilon}(u)$$

for any $p \in (0,1)$ and $\varepsilon > 0$.

Proof. Let $u_n \to u$ in $L^1(\Omega)$ and $p \in (0,1)$. Then

$$|u(x)|^p \le |u(x) - u_n(x)|^p + |u_n(x)|^p,$$

and analogously with $u(x)$ and $u_n(x)$ reversed. Consequently

$$\big||u(x)|^p - |u_n(x)|^p\big| \le |u(x) - u_n(x)|^p$$

and thus

$$|N_p(u) - N_p(u_n)| \le \int_\Omega \big||u(x)|^p - |u_n(x)|^p\big|\, dx \le \int_\Omega |u(x) - u_n(x)|^p\, dx.$$

By the Hölder's inequality

$$\int_\Omega |u(x) - u_n(x)|^p\, dx \le \int_\Omega |u(x) - u_n(x)|\, dx\, |\Omega|^{1-p} \to 0 \text{ for } n \to \infty,$$

and thus the first assertion follows. To verify the second we define the sets

$$\Omega_1 = \{|u| \ge \varepsilon^2 \ \text{ and } \ |u_n| \ge \varepsilon^2\}, \quad \Omega_3 = \{|u| < \varepsilon^2 \text{ and } |u_n| < \varepsilon^2\},$$

$$\Omega_2 = \{|u| \ge \varepsilon^2 \ \text{ and } \ |u_n| < \varepsilon^2\} \cup \{|u| < \varepsilon^2, \ \text{ and } \ |u_n| \ge \varepsilon^2\}.$$

Then, using that $|\psi_\varepsilon(t^2) \geq t^p|$ on $[0,\varepsilon]$ we find

$$|N_{p,\varepsilon}(u) - N_{p,\varepsilon}(u_n)| \leq \int_{\Omega_1} |u-u_n|^p \, d\omega + \int_{\Omega_2} |u-u_n|^p \, d\omega + \tfrac{p\varepsilon^2}{\varepsilon^{2-p}} \int_{\Omega_3} |u-u_n| \, d\omega,$$

which tends 0 for $n \to \infty$ and verifies the second claim.

□

Moreover using Lemma 6.1 and the fact that $\int_{\{x:|u_\varepsilon|<\varepsilon^2\}} |\, \frac{p}{2} \frac{|u_\varepsilon|^2}{\varepsilon^{2-p}} + (1-\frac{p}{2})\varepsilon^p| \, dx \to 0$ for $\varepsilon \to 0$, we show that

$$u_\varepsilon \to u \in L^1(\Omega) \quad \text{implies that } N_{p,\varepsilon}(u_\varepsilon) \to N_p(u). \tag{6.3}$$

Proposition 6.1. *For any $p \in (0,1)$ problem ($P_{s,\varepsilon}$) admits a solution u_ε.*

Proof. For the sake of completeness we provide proof which relies on standard techniques. Let $\{u_k\}$ denote a minimizing sequence. Then $\{|\nabla u_k|_{L^2}\}$ and $N_{p,\varepsilon}(u_k)$ are bounded sequences. Decompose any $u \in L^2(\Omega)$ as $u = u^0 + u^1$, where $u^0 = \frac{1}{\|\Omega\|} \int_\Omega u \, dx$ is a constant function, and $\int_\Omega u^1 \, dx = 0$.

Since $\{\nabla u_k\} = \{\nabla u_k^1\}$ is bounded in $L^2(\Omega)$ it follows that $\{u_k^1\}$ is bounded in $L^2(\Omega)$. This implies that $\{N_{p,\varepsilon}(u_k^1)\}$ is bounded. Since, using (2.1) and (6.2),

$$N_p(u_k^0) = N_p(u_k - u_k^1) \leq N_p(u_k) + N_p(u_k^1) \leq N_{p,\varepsilon}(u_k) + N_{p,\varepsilon}(u_k^1), \tag{6.4}$$

it follows that $\{N_p(u_k^0)\}$ is bounded as well. Hence $\{u^k\}$ is bounded in $H^1(\Omega)$. Thus there exists a subsequence, for which no new notation is introduced, and $u_\varepsilon \in H^1(\Omega)$ such that $u_k \rightharpoonup u_\varepsilon$ weakly in $H^1(\Omega)$ and $u_k \to u_\varepsilon$ strongly in $L^2(\Omega)$. As a consequence $u_\varepsilon \in \mathcal{C}$ and there exists a further subsequence, again denoted by u_k, such that $u_k \to u_\varepsilon$ almost everywhere.

By the second claim in Lemma 6.1 and using weak lower semi-continuity of norms it follows that u_ε is a solution.

□

Proposition 6.2. *Any weak accumulation point u^* in $H^1(\Omega)$ of solutions $\{u_\varepsilon\}_{\varepsilon>0}$ to ($P_{s,\varepsilon}$) as $\varepsilon \to 0$, is a solution to (P_s).*

Proof. Since $\{\nabla u_\varepsilon\}_{\varepsilon>0}$ is bounded in $L^2(\Omega)$, and $\{N_{p,\varepsilon}(u_\varepsilon)\}_{\varepsilon>0}$ is bounded, one can argue as in the previous proof that $\{u_\varepsilon\}_{\varepsilon>0}$ is bounded in $H^1(\Omega)$, and hence there exists a subsequential weak limit u^* in $H^1(\Omega)$. Using (6.3) and weak lower semi-continuity one can pass to the limit in $\mathcal{J}_\varepsilon(u_\varepsilon) \leq \mathcal{J}_\varepsilon(u)$ to obtain that $\mathcal{J}(u^*) \leq \mathcal{J}(u)$ for all $u \in \mathcal{C}$, as desired. □

Henceforth we shall use the special choice for j given by

$$j(u) = \frac{1}{2}|Ku - f|_Y^2, \tag{6.5}$$

where Y is a Hilbert space, $f \in Y$, and $K \in \mathcal{L}(H^1(\Omega), Y)$. Referring back to (3.24) the choice $K = -E^{-1}B$, $f = a - E^{-1}g$ and $Y = L^2(\Omega)$ can be considered as special case in (6.5).

The necessary optimality condition for ($P_{s,\varepsilon}$) is given by

$$-\alpha\Delta u + K^*Ku + \frac{\beta p}{\max(\varepsilon^{2-p}, |u|^{2-p})}\, u = K^*f, \tag{6.6}$$

where $|u| = |u(x)|$. To solve (6.6) we can use an iterative scheme and, given u_k, determine u_{k+1} from

$$-\alpha\Delta u_{k+1} + K^*Ku_{k+1} + \frac{\beta p}{\max(\varepsilon^{2-p}, |u_k|^{2-p})}\, u_{k+1} = K^*f. \tag{6.7}$$

By the Lax-Milgram theorem (6.7) admits a unique solution.

Theorem 6.1. *The sequence $\{\mathcal{J}_\varepsilon(u_k)\}$ generated by (6.7) is strictly decreasing and*

$$\begin{aligned} &\frac{1}{2}\sum_{k=0}^{\infty}\left(\alpha|\nabla(u_{k+1}-u_k)|_{L^2}^2 + |K(u_{k+1}-u_k)|_Y^2\right) \\ &\qquad + \frac{\beta p}{2}\sum_{k=0}^{\infty}\int_\Omega \frac{1}{\max(\varepsilon^{2-p}, |u_k|^{2-p})}|u_{k+1}-u_k|^2\,dx \le J_\varepsilon(u_0). \end{aligned} \tag{6.8}$$

Moreover there exists a subsequence of $\{u_k\}$ and some $u_\varepsilon \in H^1(\Omega)$ such that $(u_{k_l}, u_{k_{l+1}}) \rightharpoonup (u_\varepsilon, u_\varepsilon)$ in $H^1(\Omega) \times H^1(\Omega)$ and u_ε is a solution to (6.6).

Proof. Taking the inner product of (6.7) with $u_{k+1} - u_k$, for $k = 0, 1, \dots$ we have

$$\begin{aligned} &\alpha\big(\nabla u_{k+1}, \nabla(u_{k+1}-u_k)\big)_{L^2(\Omega)} + \big(Ku_{k+1}, K(u_{k+1}-u_k)\big)_Y \\ &\qquad + \Big(\frac{\beta p}{\max(\varepsilon^{2-p}, |u_k|^{2-p})}u_{k+1}, u_{k+1}-u_k\Big)_{L^2} = \langle K^*f, u_{k+1}-u_k\rangle_{(H^1)^*, H^1}, \end{aligned}$$

and hence

$$\alpha\big(\nabla u_{k+1}, \nabla(u_{k+1}-u_k)\big)_{L^2(\Omega)} + \big(Ku_{k+1}-f, K(u_{k+1}-u_k)\big)_Y$$

$$+ \big(\frac{\beta p}{\max(\varepsilon^{2-p}, |u_k|^{2-p})} u_{k+1}, u_{k+1}-u_k\big)_{L^2(\Omega)} = 0.$$

Using $b(b-a) = \frac{1}{2}(b^2 - a^2 + (a-b)^2)$ we obtain

$$\alpha|\nabla u_{k+1}|^2_{L^2(\Omega)} - \alpha|\nabla u_k|^2_{L^2} + \alpha|\nabla u_{k+1} - \nabla u_k|^2_{L^2(\Omega)} + |Ku_{k+1}-f|^2_Y - |Ku_k - f|^2_Y$$

$$+ |K(u_{k+1}-u_k)|^2_Y + \beta p \int_\Omega \frac{1}{\max(\varepsilon^{2-p}, |u_k|^{2-p})} |u_{k+1}-u_k|^2\, dx + G = 0,$$

where

$$G = \beta p \int_\Omega \frac{1}{\max(\varepsilon^{2-p}, |u_k|^{2-p})} (|u_{k+1}|^2 - |u_k|^2)\, dx$$

$$= 2\beta \int_\Omega \psi'_\varepsilon(|u_k|^2)(|u_{k+1}|^2 - |u_k|^2)\, dx \geq 2\beta \int_\Omega \big(\psi_\varepsilon(|u_{k+1}|^2) - \psi_\varepsilon(|u_k|^2)\big)\, dx.$$

Combining these estimates we arrive at

$$\alpha|\nabla u_{k+1}|^2_{L^2} + \alpha|\nabla(u_{k+1}-u_k)|^2_{L^2} + |Ku_{k+1}-f|^2_Y + |K(u_{k+1}-u_k)|^2_Y$$

$$+ 2\beta\, N_{p,\varepsilon}(u_{k+1}) + \beta p \int_\Omega \frac{1}{\max(\varepsilon^{2-p}, |u_k|^{2-p})} |u_{k+1}-u_k|^2\, dx$$

$$\leq \alpha|\nabla u_k|^2_{L^2[\Omega)} + |Ku_n - f|^2_Y + 2\beta\, N_{p,\varepsilon}(u_{k+1}),$$

and thus

$$\mathcal{J}_\varepsilon(u_{k+1}) + \frac{\alpha}{2}|\nabla(u_{k+1}-u_k)|^2 + \frac{1}{2}|K(u_{k+1}-u_k)|^2_Y$$

$$+ \frac{\beta p}{2} \int_\Omega \frac{1}{\max(\varepsilon^{2-p}, |u_k|^{2-p})} |u_{k+1}-u_k|^2\, dx \leq \mathcal{J}_\varepsilon(u_k).$$

This implies that for every k

$$\mathcal{J}_\varepsilon(u_{k+1}) + \frac{1}{2}\sum_{i=0}^{k} \big(\alpha|\nabla(u_{i+1}-u_i)|^2_{L^2} + |K(u_{i+1}-u_i)|^2_Y\big)$$
$$+ \frac{\beta p}{2}\sum_{i=0}^{k} \int_\Omega \frac{1}{\max(\varepsilon^{2-p}, |u_i|^{2-p})} |u_{i+1}-u_i|^2\, dx \leq \mathcal{J}_\varepsilon(u_0). \tag{6.9}$$

and estimate (6.8) follows. From (6.9) it follows that $k \to \mathcal{J}_\varepsilon(u_k)$ is strictly decaying, unless two consecutive elements of the sequence coincide. In this case u_k is a solution to (6.6).

From (6.9) we deduce that $\{|\nabla u_k|_{L^2}\}$ and $\{N_{p,\varepsilon}(|u_k|)\}$ are bounded sequences. Since $N(u_k) \leq N_{p,\varepsilon}(|u_k|^2) + |\Omega|\varepsilon^{2p}$ it follows that $\{N(u_k)\}$ is bounded as well. Decomposing $u_k = u_k^0 + u_k^1$ as in the proof of Proposition 6.1, we find that the constant parts $\{u_k^0\}$ are bounded, compare (6.4), and hence $\{u_k\}$ is bounded in $H^1(\Omega)$.

Hence there exists a subsequence $\{u_{k_l}\}$ and u_ε in $H^1(\Omega)$ such that $u_{k_l} \rightharpoonup u_\varepsilon$ in $H^1(\Omega)$ and $u_{k_l} \to u_\varepsilon$ strongly in $L^2(\Omega)$. Then by (6.8) we have that $u_{k_l+1} \to u_\varepsilon$ in $L^2(\Omega)$ as well. There exists a further subsequence, denoted by the same symbol, such that $(u_{k_l}, u_{k_l+1}) \to (u_\varepsilon, u_\varepsilon)$ a.e. in Ω. Passing to the limit in (6.7) we find that u_ε is a solution to (6.6). □

7 Concluding remarks

Optimization problem involving L^p regularization terms, with $p \in [0, 1)$ were investigated with special emphasis given on optimal control problems with $p = 0$. The focus in this paper was put on well-posedness of the variational problems and on obtaining a framework for optimality conditions. A detailed analysis of algorithmic aspects beyond those given on in Sections 3.4 and 6 is planned for the future. The effect of control constraints compared to $L^2(\Omega)$ control costs will be investigated. It may also be of interest to consider control costs e.g. of the form $\sum N^0(u - \bar{u}_i)$, with $\bar{u}_i$ given, in order to enhance "multi-bang" solutions.

References

[1] K. Bredies and D. Lorenz. *Mathematische Bildverarbeitung: Einführung in Grundlagen und moderne Theorie.* Vieweg und Teubner Verlag, 2010.

[2] K. Bredies and D. A. Lorenz. Minimization of non-smooth, non-convex functionals by iterative thresholding. 2009. Preprint.

[3] H. Brezis. *Operateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert.* North-Holland Pub. Co, 1973.

[4] E. Casas, R. Herzog, and G. Wachsmuth. Optimality conditions and error analysis of semilinear elliptic control problems with l1 cost functional. *SIAM Journal on Optimization.* To appear.

[5] E. Casas and J. Yong. Maximum principle for state-constrained optimal control problems governed by quasilinear elliptic equations. *Differential Integral Equation*, 8:1–19, 1995.

[6] C. Clason and K. Kunisch. A duality-based approach to elliptic control problems in non-reflexive Banach spaces. *ESAIM: Control, Optimisation and Calculus of Variations*, 17(1):243–266, 2011.

[7] I. Daubechies, M. Defrise, and C. D. Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57:1413–1457, 2004.

[8] M. Day. The spaces L^p with $0 < p < 1$. *Bull. Amer. Math. Soc.*, 46:816–823, 1940.

[9] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52:1289–1306, 2006.

[10] M. Grasmair, M. Haltmeier, and O. Scherzer. Sparse regularization with ℓ^q penalty term. *Inverse Problems*, 24(5):55020–55033, 2008.

[11] G.Stadler. Elliptic optimal control problems with L1-control cost and applications for the placement of control devices. *Computational Optimization and Applications*, 44:159–181, 2007.

[12] M. Hintermüller and T. Wu. Nonconvex TV^q-models in image restoration: Analysis and a trust-region regularization based superlinearly convergent solver. 2012. Preprint.

[13] J. Huang and D. Mumford. Statistics of natural images and models. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR1999)*, pages 541–547, 1999.

[14] K. Ito and K. Kunisch. A variational approach to sparsity optimization based on Lagrange multiplier theory. Submitted, 2011.

[15] B. Jin, T. Khan, and P. Maass. A reconstruction algorithm for electrical impedance tomography based on sparsity regularization. *Int. Journal on Numerical Mathematics and Engingeering*, 2011. Published online.

[16] T. P. N. Kalton and J. Roberts. *An F-space Sampler*, volume 89 of *London Mathematical Society Lecture Notes Series.* Cambridge University Press, 1984.

[17] M. Nikolova. Minimizers of cost-functions involving nonsmooth data-fidelity terms. applica- tion to the processing of outliers. *SIAM Journal on Numerical Analysis*, 40:965–994, 2002.

[18] C. A. Zarzer. On Tikhonov regularization with non-convex sparsity constraints. *Inverse Problems*, 25:25006–, 2009.