SIAM J. CONTROL OPTIM.
Vol. 52, No. 2, pp. 1251–1275

OPTIMAL CONTROL WITH $L^p(\Omega)$, $p \in [0,1)$, CONTROL COST*

KAZUFUMI ITO† AND KARL KUNISCH‡

Abstract. L^p optimal control with $p \in [0,1)$ is investigated. The difficulty of natural lack of convexity and thus 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 with quadratic tracking and regularized L^0 control costs are given. 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.

Key words. L^0 minimization, optimal control, bang-bang control, sparsity optimization, maximum principle, nonsmooth optimization, primal-dual active set method

AMS subject classifications. 49J20, 49J30, 49K20, 49J52

DOI. 10.1137/120896529

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 and subsequent numerical treatment. The use of higher order polynomial powers, corresponding to L^p-norms, $p > 1$, were motivated, for instance, in optimal control of semilinear 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 [4, 10, 9, 17] and the references there for imaging and [6, 8, 13] 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 [16], 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 [3, 12, 14, 19, 22], for example. As we shall see the use of L^p, $p \in [0,1)$, norms offers interesting applications for sparse controls and controls of bang-bang-bang type.

Let us describe the contents of the paper. In section 2 we briefly recall some properties of the vector space $L^0(\Omega)$. It will then be obvious that the use of the L^0-functional calls for special techniques to overcome the lack of convexity and weak lower semicontinuity. This also applies for L^p with $p \in [0,1)$. In our work we endow L^0 with the Ekeland metric. Its use in the cost functional implies a volume constraint.

*Received by the editors October 25, 2012; accepted for publication (in revised form) December 2, 2013; published electronically April 24, 2014.
http://www.siam.org/journals/sicon/52-2/89652.html

†Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205 (kito@math.ncsu.edu). This author's research was partially supported by the Army Research Office under DAAD19-02-1-0394.

‡Institut für Mathematik und wissenscahftliches Rechnen, Karl-Franzens-Universität Graz, A-8010 Graz, Austria (karl.kunisch@uni-graz.at). This author's research was supported in part by the Fonds zur Förderung der wissenschaftlichen Forschung under SFB 32, Mathematical Optimization and Applications in the Biomedical Sciences.

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 [10].

As mentioned above, to establish existence for variational problems involving L^p-functionals with $p \in [0,1)$, one cannot rely on convexity and weak lower semicontinuity properties. In the case of ℓ^p an appropriately chosen transformation can be applied, and the fact that the duality mapping from $\ell^p \to \ell^{p'}$, with $\frac{1}{p}+\frac{1}{p'}=1$, is weakly sequentially continuous [16, 22] can be used to ascertain existence. We propose to meet the difficulties involved in the L^p-case, $p \in [0,1)$, by appropriately chosen regularization terms or norm bounds. The choice of the regularization terms depends 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 necessarily allow such a splitting. The first case is treated by regularization with L^2-functionals in section 2 and alternatively by bilateral pointwise bounds in section 3. For the latter case regularization in H^1 is used and analyzed in sections 4 and 5. For L^2 regularization or pointwise norm bounds, the solutions are possibly only relaxed controls, where the effect of relaxation can be quantified. In the case of H^1 regularization, existence can be guaranteed for cost functionals which contain nonconvex L^p terms without the need for relaxation. The use of regularization or relaxation for nonconvex problems to guarantee existence of the variational problem itself and/or the associated optimality condition is classical. From the vast literature we mention [11, 18, 21]. While starting from a different perspective, our approach is closely related to that described for a general class of nonconvex problems in [11]; see Remark 2.10.

Sections 2 and 3 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 a linear state equation. 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 process leads to a relaxed control. We will derive an estimate on the error in the cost corresponding to the relaxed control and the infimum of the optimal control problem. This approach is first carried out for problems without constraints on the controls in section 2, where also the first step toward numerical realization on the basis of a primal dual active set strategy is proposed. In section 3 the case of bilateral constraints on the controls 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; they are therefore bang-bang-bang type.

In section 4 we consider the existence of 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 5 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 [15]. Subsequently we analyze convergence of a monotone scheme to solve the L^p-problems iteratively. A brief outlook concludes the paper.

2. $L^0(\Omega)$ optimal control.

This section is devoted to optimal control problems involving the functional

$$N_0(f) = \int_\Omega |f(x)|^0 \, dx,$$

where we set $0^0 = 0$. It is related to $L^0(\Omega)$, 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, and we have $N_0(f) = \text{meas}\{x \in \Omega : f(x) \neq 0\} = d_0(f, 0)$. We note that $f \to N_0(f)$ is subadditive, but it is not positive homogeneous. Further, if f and g are different from 0 a.e., then for any $\lambda \in [0, 1]$ we have

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

but this inequality is not true for $f \neq 0$ and $g = 0$ and hence $f \to N_0(f)$ is not convex.

2.1. Necessary optimality.

From the above considerations it is obvious that standard existence results are not applicable for variational problems posed in L^0. Here we proceed differently and start by deriving necessary optimality conditions for a general purpose functional with respect to the cost u, assuming the existence of a solution. These results in particular are applicable for $N_0(u)$. Subsequently these necessary conditions are investigated with respect to existence and the correspondence of these solutions to the original variational problem. This analysis focuses on the N_0-functional combined with a quadratic one.

We start by considering a class of 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.

Let X be a real Hilbert space that is densely and compactly embedded in to $L^2(\Omega)$ such that $X \subset L^2(\Omega) \subset X^*$ are a Gelfand triple. Here Ω is a bounded domain in $\mathbb{R}^n$ which describes the (time-) space domain of the control system and is endowed with the Lebesgue measure. The state variable and the control variable are denoted by x and u, respectively.

We consider the constrained minimization problem

$$\min \quad \int_\Omega \big(\ell(\omega, x(\omega)) + h(u(\omega))\big) \, d\omega \tag{2.1}$$

subject to the equality constraint

$$E\, x + f(\cdot, x, u) = 0 \quad \text{in } X^* \tag{2.2}$$

over

$$u \in U_{ad} = \{u \in L^2(\Omega) : u(\omega) \in U \ \text{ a.e}\}. \tag{2.3}$$

Here U is a closed convex subset of $\mathbb{R}$ and $E \in \mathcal{L}(X, X^*)$ with X^* the dual space to X. Further, $f \in C^1(\Omega \times \mathbb{R}^2, \mathbb{R})$, $\ell : \Omega \times \mathbb{R} \to \mathbb{R}$ is measurable and C^1 with respect to the second variable, and $h : \mathbb{R} \to \mathbb{R}$ is measurable. The mappings f, ℓ, h give rise to substitution operators which are denoted by the same symbols and are supposed to satisfy

$$\text{(2.4)} \qquad \begin{cases} (x,u) \in X \times U_{ad} \to f(\cdot,x,u) \in L^2(\Omega), \\ x \in X \to \ell(\cdot,x) \in L^1(\Omega), \\ u \in U_{ad} \to h(u) \in L^1(\Omega). \end{cases}$$

Throughout it is assumed that (2.2) 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 (2.1)–(2.3) 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:

$$\text{(2.5)} \qquad \begin{cases} (x,u) \in X \times U_{ad} \to f_x(\cdot,x,u) \in \mathcal{L}(X,X^*), \\ \ell_x(\cdot,x) \in L^1(\Omega) \cap X^* \text{ for } x \in X. \end{cases}$$

With (2.5) holding the following adjoint equation is well-defined:

$$\text{(2.6)} \qquad (E + f_x(\cdot,\bar{x},\bar{u}))^* p + \ell_x(\cdot,\bar{x}) = 0.$$

It is assumed that (2.6) admits a unique solution $p \in X$.

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

$$\text{(2.7)} \qquad v(\omega) = \begin{cases} u & \text{on } B(s,\delta) = \{\omega : |\omega - s| < \delta\}, \\ \bar{u}(\omega) & \text{otherwise,} \end{cases}$$

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 [2, 7].

We denote by $x = x(v)$ the corresponding solution of (2.2). The following additional properties for the optimal state $\bar{x}$ and each perturbed state v will be needed:

$$\text{(2.8)} \quad \begin{cases} |x(v) - \bar{x}|^2_{L^2(\Omega)} = o\,(\mathrm{meas}(B(s,\delta))), \\ \int_\Omega \big(\, \ell(\cdot,x(v)) - l(\cdot,\bar{x}) - \ell_x(\cdot,\bar{x})(x(v)-\bar{x})\big) d\omega \; = O(|x(v)-\bar{x}|^2_{L^2}), \\ \langle f(\cdot,x(v),v) - f(\cdot,\bar{x},v) - f_x(\cdot,\bar{x},v)(x(v)-\bar{x}), p\rangle_{X^*,X} \; = O(|x(v)-\bar{x}|^2), \\ (f_x(\cdot,\bar{x},u) - f_x(\cdot,\bar{x},\bar{u}))p \in L^2(\Omega) \text{ for each } u \in U. \end{cases}$$

Remark 2.1. The first assumption in (2.8) is well-established in the context of ordinary differential equations. 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 $B \in \mathcal{L}(U, X^* \cap L^1(\Omega))$ 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

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

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 (2.7). Then by the above estimates

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

which implies the first estimate in (2.8). For E an elliptic operator with sufficiently smooth coefficients we have

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

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

THEOREM 2.2. *Suppose* $(\bar{x}, \bar{u}) \in X \times U_{ad}$ *is optimal for problem* (2.1), *that* $p \in X$ *satisfies the adjoint equation* (2.6), *and that* (2.4), (2.5), *and* (2.8) *hold. Then we have the necessary optimality condition: for each* $u \in U$

$$\mathcal{H}(\omega, \bar{x}(\omega), u, p(\omega)) - \mathcal{H}(\omega, \bar{x}(\omega), \bar{u}(\omega), p(\omega)) \geq 0 \text{ for a.e. } \omega \in \Omega. \tag{2.9}$$

Proof. By the second property in (2.8) 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 (2.7) and $x = x(v)$. Utilizing the adjoint equation (2.6) we find that

$$\begin{aligned} &0 \leq J(v) - J(\bar{u}) \\ &\quad = -\langle (E + f_x(\cdot, \bar{x}, \bar{u}))(x - \bar{x}), p)\rangle_{X^*,X} + \int_\Omega \big(h(v) - h(\bar{u})\big)\, d\omega + O(|x - \bar{x}|^2). \end{aligned} \tag{2.10}$$

By the third property in (2.8) we have

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

By (2.4), (2.10), and the fourth property in (2.8) we find

$$
\begin{aligned}
0 &\le J(u) - J(\bar u)\\
&= \int_\Omega (f(\cdot,\bar x,v) - f(\cdot,\bar x,\bar u))p\,d\omega + \int_\Omega (h(v) - h(\bar u))\,d\omega\\
&\quad + \int_\Omega (f_x(\cdot,\bar x,v) - f_x(\cdot,\bar x,\bar u))(x-\bar x)\,p\,d\omega + O(|x-\bar x|^2)\\
&\le \int_\Omega (f(\cdot,\bar x,v) - f(\cdot,\bar x,\bar u))p\,d\omega + \int_\Omega (h(v) - h(\bar u))\,d\omega\\
&\quad + \left(\int_\Omega |f_x(\cdot,\bar x,v)\,p - f_x(\cdot,\bar x,\bar u)\,p|^2\,d\omega\right)^{\frac12}\ |x-\bar x|_{L^2} + O(|x-\bar x|^2).
\end{aligned}
$$

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

$$
\begin{aligned}
\omega \to \big(&h(\bar u)(\omega),\ (f(\omega,\bar x(\omega),u) - f(\omega,\bar x(\omega),\bar u(\omega)))p(\omega),\\
&|\big(f_x(\omega,\bar x(\omega),u) - f_x(\omega,\bar x(\omega),\bar u(\omega))\big)\,p(\omega)|^2\big).
\end{aligned}
\tag{2.11}
$$

Let $S = S(u)$ denote the set of these Lebesgue points and note that $\operatorname{meas}(S(u)) = \operatorname{meas}(\Omega)$. Dividing the above string of inequalities by $|B(s,\delta)| > 0$ for arbitrary $s \in S(u)$, letting $\delta \to 0$, and using the first property in (2.8) we obtain

$$
\mathcal{H}(s,\bar x(s),u,p(s)) - \mathcal{H}(s,\bar x(s),\bar u,p(s)) \ge 0 \text{ for } s \in S(u),
\tag{2.12}
$$

and the claim follows. ∎

COROLLARY 2.3. *If, under the assumption of Theorem* 2.2, *in addition there exists a Lebesgue measurable set* Ω_0 *with* $meas(\Omega_0) = meas(\Omega)$ *such that for all* $u \in U$ *the set of Lebesgue points of*

$$
\omega \to \big(f(\omega,\bar x(\omega),u)p(\omega),\ |(f_x(\omega,\bar x(\omega),u) - f_x(\omega,\bar x(\omega),\bar u(\omega)))\,p(\omega)|^2\big)
$$

contains Ω_0, *then*

$$
\mathcal{H}(\omega,\bar x(\omega),u,p(\omega)) - \mathcal{H}(\omega,\bar x(\omega),\bar u(\omega),p(\omega)) \ge 0 \text{ for a.e. } \omega\in\Omega \text{ and all } u \in U.
\tag{2.13}
$$

Proof. We have $\operatorname{meas}(\Omega_1 \cap \Omega_0) = \operatorname{meas}(\Omega)$, where Ω_1 is the set of all Lebesgue points of $\omega \to \big(h(\bar u)(\omega),\ f(\omega,\bar x(\omega),\bar u(\omega))p(\omega)\big)$. The assertion now follows with (2.11) and (2.12). ∎

If $f(\cdot,x,u)$ is of the form $f(\cdot,x,u) = \tilde f(\cdot,x) + Bu$ with $B \in \mathcal{L}(U, L^2(\Omega)$, then for constant functions $u \in U$ we have $B(u) = B(u\,1) = uB(1)$, where 1 is the constant function with value 1 and hence (2.13) is automatically satisfied.

Remark 2.4. Let us observe that the condition involving the Lebesgue points in the previous corollary is satisfied if for every $\epsilon > 0$ there exists $\rho > 0$ such that for almost all $\omega \in \Omega$ and all $u, v \in U$

$$
|f_x(\omega,\bar x(\omega),u) - f_x(\omega,\bar x(\omega),v)| < \epsilon \text{ if } |u-v| < \rho
$$

holds. Indeed, since $U \subset \mathbb{R}$ there exists a sequence $\{u_i\}_{i=1}^\infty$ which is dense in U. Let Ω_i be the set of Lebesgue points of

$$
\omega \to \big(f(\omega,\bar x(\omega),u_i)p(\omega),\ |(f_x(\omega,\bar x(\omega),u_i) - f_x(\omega,\bar x(\omega),\bar u(\omega)))\,p(\omega)|^2\big).
$$

Then $\Omega_0 = \cap_{i=1}^\infty \Omega_i$ satisfies the assumption of the corollary.

2.2. 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 (2.1)–(2.3) and the necessary condition (2.9) in the case that h involves the N_0-functional.

More specifically we consider the case of (2.1) when $h : \mathbb{R} \to \mathbb{R}$ is given by

$$h(u) = \frac{\alpha}{2} u^2 + \beta |u|_0, \tag{2.14}$$

where

$$|u|_0 = \begin{cases} 0 \text{ if } u = 0, \\ 1 \text{ if } u \neq 0, \end{cases}$$

and α and β are positive constants. The resulting optimal control problem is then given by

$$\begin{cases} \quad \min & \int_\Omega (\ell(\cdot, x) + \frac{\alpha}{2}|u|^2 + \beta |u|_0)\, d\omega, \\ \text{subject to} & Ex + f(x) + Bu = 0, \quad u \in L^2(\Omega), \end{cases} \tag{2.15}$$

where $E \in \mathcal{L}(X, X^*)$, $B \in \mathcal{L}(L^2(\Omega))$, and ℓ, f satisfy (2.4) and (2.5). The case of control constraints will be considered separately in section 3 below.

The maximum principle established in Theorem 2.2 suggests considering $\operatorname{argmin}_{u \in \mathbb{R}}(h(u) + qu)$. A short computation then shows that

$$\Phi(q) := \operatorname*{argmin}_{u \in \mathbb{R}} (h(u) + qu) = \begin{cases} -\frac{q}{\alpha} & \text{for } |q| \geq \sqrt{2\alpha\beta}, \\ 0 & \text{for } |q| < \sqrt{2\alpha\beta}. \end{cases} \tag{2.16}$$

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

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

The mapping $-\tilde{\Phi} : \mathbb{R} \to 2^{\mathbb{R}}$ is maximal monotone. We also note that

$$h(\tilde{\Phi}(q)) + q\,\tilde{\Phi}(q) = [0, \beta) \text{ for } q = |\sqrt{2\alpha\beta}|, \tag{2.17}$$

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 2.3.

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.

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

$$\begin{cases} Ex + f(x) + B\Phi(B^* p) = 0, \\ (E + f'(x))^* p + \ell'(\cdot, x) = 0, \end{cases} \tag{2.18}$$

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

$$F(u) - F(u^*) \geq (B^* p, u - u^*)_{L^2(\Omega)} \quad \textit{for all } u \in L^2(\Omega), \tag{2.19}$$

then u^* *is a solution to* (2.15).

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

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

We note that the expressions on the right-hand side of the above identity are nonnegative. Together with (2.19) this gives the desired result. ∎

We next turn to discuss conditions (2.18) and (2.19). For this purpose we consider the special case of a linear state equation

$$Ex + Bu = g \qquad (\text{i.e., } f(x, u) = Bu - g) \tag{2.20}$$

with $g \in X$ and assume that

$$x \to \ell(\omega, x) \text{ is convex.} \tag{2.21}$$

From the adjoint equation and

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

it follows that

$$F(u) - F(u^*) = (B^* p, u - u^*)_{L^2(\Omega)} + \int_\Omega (\ell(\cdot, x) - \ell(\cdot, x^*) - \ell'(x^*)(x - x^*))\, d\omega, \tag{2.22}$$

and hence the convexity assumption (2.21) for ℓ implies (2.19). If in addition to (2.20) we assume that $\ell_x = x - a$ with $a \in X$, then (2.18) 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 therefore introduce a relaxation by replacing the monotone operator $-\Phi$ by its maximal monotone extension $-\tilde{\Phi}$. This results in the system

$$\begin{cases} Ex + B\tilde{\Phi}(B^*p) \ni g, \\ E^*p + x = a. \end{cases} \tag{2.23}$$

PROPOSITION 2.6. *If $E \in \mathcal{L}(X, X^*)$ is an isomorphism, then* (2.23) *admits a unique solution $(x,p) \in X \times X$ and there exists a constant K such that*

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

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., [20, p. 19]. Further define

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

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 define $\Psi : \mathbb{R} \to \mathbb{R}$ by

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

and note that Ψ is a proper convex function with $\partial(\Psi) = -\tilde{\Phi}$. Hence $-\tilde{\Phi}$ is maximal monotone [5, p. 24] and the associated substitution operator (denoted by the same symbol) from $L^2(\Omega)$ to $L^2(\Omega)$ is maximal monotone as well [5, p. 25]. Since the domain of B^* is all of $L^2(\Omega)$ it follows that $-B\tilde{\Phi}B^*\cdot$ is maximal monotone from $L^2(\Omega)$ to $L^2(\Omega)$ as well [1, Theorem 24.5]. Finally, since the domain of $B\tilde{\Phi}B^*$ is all of $L^2(\Omega)$ it follows that $EE^* - B\tilde{\Phi}B^*$ is a maximal monotone in $L^2(\Omega)$; see, e.g., [5, Corollary 2.7].

For $a \in D(E)$ consider the equation in $L^2(\Omega)$

$$EE^*p - B\tilde{\Phi}B^*p \ni Ea - g. \tag{2.24}$$

Since

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

it follows that $EE^* - B\tilde{\Phi}B^*$ is coercive and hence (2.24) admits a solution $p \in D(EE^*)$; see, e.g., [5, p. 31]. It satisfies $|E^*p|^2_{L^2(\Omega)} \le (Ea+g,p)_{L^2(\Omega)}$ and hence

$$|E^*p|_{L^2(\Omega)} \le |a|_{L^2(\Omega)} + |E^{-1}g|_{L^2(\Omega)} \le |a|_{L^2(\Omega)} + K|g|_{L^2(\Omega)}.$$

Here and below K denotes a constant independent of (a,g). With p determined, define x by $x = a - E^*p \in D(E)$. The pair (x,p) satisfies (2.23). By (2.24) we have $Ex = Ea - EE^*p \in g - B\tilde{\Phi}B^*p$ and hence K can be chosen such that $|Ex|_{L^2(\Omega)} \le K(|a|_{L^2(\Omega)} + |g|_{L^2(\Omega)})$ and, since $E \in \mathcal{L}(E,E^*)$ is an isomorphism, we have

$$|(x,p)|_{X\times X} \le K(|a|_{L^2(\Omega)} + |g|_{L^2(\Omega)}). \tag{2.25}$$

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

$$\begin{cases} Ex_n + y_n = g \quad \text{with } y_n \in B\tilde{\Phi}B^*p_n, \\ E^*p_n + x_n = a_n. \end{cases} \tag{2.26}$$

By (2.25) 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, the closedness 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 [5, p. 27], it follows that $y \in B\tilde{\Phi}B^*p$ and we can pass to the limit in (2.26) to obtain that (x,p) satisfies (2.23) and also (2.25).

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

$$\begin{aligned} &E\,\delta x + B\tilde{\Phi}(B^*p) - B\tilde{\Phi}(B^*\bar{p}) \ni 0, \\ &E^*\,\delta p + \delta x = a. \end{aligned}$$

Taking the inner product with $(\delta p, -\delta x)$ in the above equations and adding them up we obtain $|\delta x|^2_{L^2(\Omega)} \ge 0$. This implies that $\delta x = 0$ and further $\delta p = 0$. ◻

We summarize the above developments for the problem

$$\begin{cases} \min J(x,u) = \frac{1}{2}|x-z|^2_{L^2(\Omega)} + \frac{\alpha}{2}|u|^2_{L^2(\Omega)} + \beta N_0(u), \\ Ex + Bu = g, \quad u \in L^2(\Omega), \end{cases} \tag{2.27}$$

where $z \in L^2(\Omega)$, $\alpha > 0, g \in X^*$. 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 $\frac{1}{2}|x-z|^2_{L^2(\Omega_o)}$ with $\Omega_o \subsetneqq \Omega$.

THEOREM 2.7. *Consider problem* (2.27) *with $E \in \mathcal{L}(X,X^*)$ an isomorphism. Let (x^*,p) be the unique solution to* (2.23) *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* (2.27) *which satisfies the optimality system*

$$\begin{cases} Ex + Bu^* = g, \\ E^*p + x - z = 0, \\ u^* = \begin{cases} \frac{-B^*p}{\alpha} & \textit{for } |B^*p| \ge \sqrt{2\alpha\beta}, \\ 0 & \textit{for } |B^*p| < \sqrt{2\alpha\beta}, \end{cases} \end{cases} \tag{2.28}$$

with the last equality holding pointwise a.e.

Remark 2.8. If meas$\{\omega : |p(\omega)| = \sqrt{2\alpha\beta}\} > 0$, then (2.27) 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 (2.17) and will be considered further in Theorem 2.12 below. The fact that the optimal control is identically equal to zero for p sufficiently small justifies calling N_0 a sparsity enhancing functional.

Remark 2.9. Introducing $\lambda = \alpha u + B^* p$ the complementarity system in (2.28) can equivalently be expressed as

$$\begin{cases} \lambda = 0 & \text{if } |\lambda - \alpha u| > \sqrt{2\alpha\beta}, \\ u = 0 & \text{if } |\lambda - \alpha u| \le \sqrt{2\alpha\beta}. \end{cases} \tag{2.29}$$

This form should be compared to the optimality system that was obtained in [16] 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 the case of strict complementarity is given by

$$\begin{cases} \lambda_i = 0 & \text{if } \left|\lambda_i - (|A_i|^2 + \alpha)x_i^*\right| > \sqrt{2\beta}(|A_i|_2 + \sqrt{\alpha}), \\ \bar{x}_i = 0 & \text{if } \left|\lambda_i - (|A_i|^2 + \alpha)x_i^*\right| \le \sqrt{2\beta}(|A_i|_2 + \sqrt{\alpha}), \end{cases} \tag{2.30}$$

where $A_i = Ae_i$, with e_i the element in ℓ^2 which has 1 in the ith element and is 0 otherwise. Comparing to (2.29) we note that in the discrete formulation the tracking part of the cost sustains in the optimality condition, whereas in (2.29) it does not. Thus for this class of problems, the order of discretization and optimization makes a significant difference.

Remark 2.10. Proceeding formally we show that the relaxation that we used by passing from Φ to $\tilde{\Phi}$ involving the Hamiltonian can equivalently be obtained by applying the Γ-regularization for separable cost functionals as investigated in [11, Chapter 9.3], for example. This involves the bi-conjugate h^{**} of h defined in (2.14). It is given by

$$h^{**}(r) = \begin{cases} \frac{\alpha r^2}{2} + \beta & \text{for } |r| \ge \sqrt{\frac{2\beta}{\alpha}}, \\ \sqrt{2\alpha\beta}\,|r| & \text{for } |r| < \sqrt{\frac{2\beta}{\alpha}}. \end{cases}$$

Note that h^{**} is convex and it is C^1 except at the origin. The relaxed problem corresponding to (2.15) using Γ-regularization results from replacing h by h^{**} and is given by

$$\begin{cases} \min & \int_\Omega (\ell(\cdot, x) + h^{**}(u))\, d\omega, \\ \text{subject to} & Ex + f(x) + Bu = 0, \quad u \in L^2(\Omega). \end{cases} \tag{2.31}$$

Note that

$$\partial h^{**}(r) = \begin{cases} \alpha\, r & \text{for } |r| \ge \sqrt{\frac{2\beta}{\alpha}}, \\ \sqrt{2\alpha\beta}\, Sgn(r) & \text{for } |r| < \sqrt{\frac{2\beta}{\alpha}}, \end{cases}$$

where $Sgn(r) = 1$ for $r > 0$, $Sgn(r) = -1$ for $r < 0$, and $Sgn(0) = [-1, 1]$. It can be checked that that ∂h^{**} is maximal monotone with inverse given by $(\partial h^{**})^{-1}$ by $-\tilde{\Phi}$.

The necessary optimality condition for problem (2.15) is then found to be

$$\begin{cases} Ex + f(x) - B(\partial h^{**})^{-1}(B^*p) \ni g, \\ (E + f'(x))^* p + \ell'(\cdot, x) = 0. \end{cases}$$

which coincides with (2.18) after replacing Φ by its relaxation $-\tilde{\Phi}$.

2.3. Uniqueness.

THEOREM 2.11. *If $E \in \mathcal{L}(X, X^*)$ is an isomorphism and u^* is a solution to* (2.27)*, 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

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

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

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

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

$$\begin{aligned} J(x,u) - J(x^*,u^*) &= \beta \textstyle\int_{S^+} (\chi_{|u| \neq 0} - \chi_{|u^*| \neq 0}) d\omega \\ &\quad + \textstyle\int_{S^0} (qu + \beta\chi_{u \neq 0}) d\omega + \frac{\alpha}{2}|u - u^*|^2_{L^2(\Omega)} + \frac{1}{2}|x - x^*|^2_{L^2(\Omega)}. \end{aligned} \tag{2.32}$$

We further set

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

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

(2.33)

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

If $\text{meas}(S_0^0) \neq 0$, 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 $\text{meas}(S_0^+) = 0$ and $J(x, u) = J(x^*, u^*)$, then $u = u^*$ a.e. on S^+ and hence $u = u^*$ a.e. in Ω. Otherwise $\text{meas}(S_0^+) > 0$ 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 (2.27). ∎

In the following result we quantify the quality of the relaxed optimal control that is obtained from the optimality system (2.23) as a suboptimal solution to (2.27).

THEOREM 2.12. *Under the assumptions of Theorem* 2.7, *if* $u^* \in \tilde{\Phi}(B^* p)$ *with* (x^*, p) *the solution to* (2.23), *then*

$$J(x(u), u) > J(x(u^*), u^*) - \beta \, \text{meas}\, \mathcal{S} \tag{2.34}$$

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

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

where for the last estimate we can proceed as in the proof of Theorem 2.11. A simple computation and (2.17) imply that

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

and the inequality is strict, if $\text{meas}\, \mathcal{S} > 0$. This implies (2.34) with $>$ replaced by $\geq$. Strict inequality holds since it was already obtained in Theorem 2.11 that if $\text{meas}\, \mathcal{S} = 0$ and if $\text{meas}\, \mathcal{S} > 0$, then it follows from (2.17). ∎

Example 2.13. We close this section with an example which illustrates the sparsity enhancing property of the functional h given in (2.14). It also provides an example for which the set $\mathcal{S}$ of Theorem 2.12 is empty.

Consider the optimal control problem (2.27) with $X = H_0^1(\Omega)$, $\Omega = (0, 1)$, $E = -\Delta \in \mathcal{L}(H_0^1(\Omega), H^{-1}(\Omega))$, $B = I$, $\alpha = \frac{1}{4}$, $\beta = 1$, and set $z = (1 + \pi^2)\sin(\pi\omega)$, and

$$g = \begin{cases} (\pi^2 - 4)\sin(\pi\omega) & \text{for } \omega \in \Omega_1, \\ \pi^2 \sin(\pi\omega) & \text{for } \omega \in \Omega_2, \end{cases}$$

where $\Omega_1 = [\frac{1}{4}, \frac{3}{4}]$, $\Omega_2 = \Omega \setminus \Omega_1$. A direct computation shows that $x(\omega) = p(\omega) = \sin(\pi\omega)$ is a solution to (2.23). Moreover the set $\{\omega : \sin(\pi\omega) = \sqrt{\frac{1}{2}} = 0\} = \{\frac{1}{4}, \frac{3}{4}\}$ has measure 0. We find that (x, p, u^*) with

$$u^* = \begin{cases} -4 \sin(\pi\omega) & \text{for } \omega \in \Omega_1, \\ 0 & \text{for } \omega \in \Omega_2 \end{cases}$$

is a solution of (2.28). By Theorem 2.12, $(x(u^*), u^*)$ is the unique solution to the optimal control problem. We point out its sparsity property and recall that for $\beta = 0$, with the other specifications unchanged the optimal control would be in $H^4(\Omega)$.

2.4. Primal-dual active set methods. The complementary condition (2.29) suggests a primal-dual active set strategy to solve (2.27). It does not realize the set-valued nature of (2.23), 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

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

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

- 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

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

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

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

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 $X = H_0^1(\Omega)$, Ω the unit square, and discretization based on finite differences with respect to a uniform mesh $h = 1/N$ to solve (2.27).

Example 2.14. 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 $\text{tol}_1 = 10^{-11}$, $\text{tol}_2 = \frac{1}{N}$.

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 satisfies the stopping criterion.

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 situation in which 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 (2.27) 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

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

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 the following:

- 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

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

Example 2.15. 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 the critical set, where $|p| \sim \sqrt{2\alpha\beta}$, is different. In Table 2, $|\mathcal{C}_2|$ stands for the number of nodes in $\mathcal{C}_2 = \{\sqrt{2\alpha\beta} - \epsilon \le |B^*p^n| \le \sqrt{2\alpha\beta} + \epsilon\}$.

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
$\|\mathcal{C}_2\|$	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

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

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 15,148 and 15,013. We also confirmed that in the final iteration, the control satisfies $|u| \le \frac{\sqrt{2\alpha\beta}+\epsilon}{\alpha}$ over the region where $\{\sqrt{2\alpha\beta} - \epsilon \le |p| \le \sqrt{2\alpha\beta} + \epsilon\}$ which accounts for closing the graph of Φ at $\sqrt{2\alpha\beta}$.

Example 2.16. In this example the control and observation occupy only part of the domain and they are nonoverlapping. 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.

It should be noted that for the examples presented here, the corresponding optimal controls are zero on sets which contain interior points. Thus these types of sparse controls differ from those obtained by with $L^1(\Omega)$ (or more precisely, measure valued costfunctionals). 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.

3. L^0: Optimal control with control constraints. We return to problem (2.27) but this time, rather than involving an $L^2(\Omega)$-regularization term, we utilize pointwise constraints on the controls:

$$\left\{ \begin{array}{ll} \min & \int_\Omega (\ell(\cdot, x) + \beta |u|_0)\, d\omega, \\ \text{subject to} & Ex + f(x) + B(u) = 0, \\ & u \in U_{ad} = \{u \in L^2(\Omega) : a \le u(\omega) \le b \text{ a.e.}\}, \end{array} \right. \tag{3.1}$$

where $a < 0 < b$.

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

As in section 2.2 we start with preliminaries involving the Hamiltonian. We define $h : \mathbb{R} \to \mathbb{R}$ by

$$h(u) = \beta |u|_0 + \chi_{[a,b]}, \tag{3.2}$$

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

$$\Phi(q) := \operatorname*{argmin}_{u\in\mathbb{R}} (h(u)+qu) = \begin{cases} 0 & \text{for } -\frac{\beta}{b} < q < -\frac{\beta}{a}, \\ b & \text{for } q \le -\frac{\beta}{b}, \\ a & \text{for } q \ge -\frac{\beta}{a}, \end{cases} \tag{3.3}$$

and

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

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

$$\tilde{\Phi}(q) = \begin{cases} 0 & \text{for } -\frac{\beta}{b} < q < -\frac{\beta}{a}, \\ b & \text{for } q < -\frac{\beta}{b}, \\ a & \text{for } q > -\frac{\beta}{a}, \\ [0,b] & \text{for } q = -\frac{\beta}{b}, \\ [a,0] & \text{for } q = -\frac{\beta}{a}. \end{cases} \tag{3.4}$$

We also have

$$h(\tilde{\Phi}(q)) + q\,\Phi(q) = [0,\beta) \ \text{ for } q \in \left\{-\frac{\beta}{b}, -\frac{\beta}{a}\right\}. \tag{3.5}$$

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 (2.9), the optimal controls, aside from the switching points $\{\frac{-\beta}{b}, \frac{-\beta}{a}\}$, can assume only three, respectively, two states.

THEOREM 3.1. *Consider problem* (3.1) *with* $E \in \mathcal{L}(X, X^*)$ *an isomorphism. Let* (x^*, p) *be the unique solution to* (2.23), *with* $\tilde{\Phi}$ *given in* (3.4), *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* (3.1) *admits a solution* (u^*, x^*) *which satisfies the optimality system*

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

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

Proof. The arguments of the proof of Proposition 2.6 guarantee the existence of a solution to (2.23) with $\tilde{\Phi}$ given in (3.4).

It suffices to argue that (x^*, u^*) is indeed a solution to (3.1). For this purpose we estimate

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

Using (2.19) and (2.22) the claim follows. ∎

Concerning the relaxed solution, we again have an estimate analogous to that in Theorem 2.12.

THEOREM 3.2. *Under the assumptions of Theorem* 3.1*, if* $u^* \in \tilde{\Phi}(B^*p)$ *with* (x^*, p) *the solution to* (2.23)*, then*

$$J(x(u), u) > J(x(u^*), u^*) - \beta\, \text{meas}\,\mathcal{S} \tag{3.6}$$

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 2.12 we find for $u \in U_{ad}$

$$\begin{aligned}
J(x(u), u) - J(x(u^*), u^*) &= (q, u - u^*)_{L^2(\Omega\setminus\mathcal{S})} + \beta \textstyle\int_{\Omega\setminus\mathcal{S}}(|u|_0 - |u^*|_0)\, d\omega \\
&\quad + (q, u - u^*)_{L^2(\mathcal{S})} + \beta \textstyle\int_{\mathcal{S}}(|u|_0 - |u^*|_0)\, d\omega \\
&\geq (-\tfrac{\beta}{b}, u)_{L^2(\mathcal{S}_1)} + \beta \textstyle\int_{L^2(\mathcal{S}_1)} |u|_0\, d\omega + (-\tfrac{\beta}{a}, u)_{L^2(\mathcal{S}_2)} + \beta \textstyle\int_{L^2(\mathcal{S}_2)} |u|_0\, d\omega \\
&\quad + (\tfrac{\beta}{b}, u^*)_{L^2(\mathcal{S}_1)} + (\tfrac{\beta}{a}, u^*)_{L^2(\mathcal{S}_2)} - \beta \textstyle\int_{L^2(\mathcal{S})} |u^*|_0\, d\omega,
\end{aligned}$$

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

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

For $u \neq u^*$ and $\text{meas}\,\mathcal{S} < \text{meas}\,\Omega$ this inequality is strict, as already noted above. For $\text{meas}\,\mathcal{S} > 0$ we have $(\frac{\beta}{b}, u^*)_{L^2(\mathcal{S}_1)} + (\frac{\beta}{a}, u^*)_{L^2(\mathcal{S}_2)} - \beta \int_{L^2(\mathcal{S})} |u^*|_0\, d\omega > -\beta\, \text{meas}\,\mathcal{S}$ by (3.5). This concludes the proof. ∎

Remark 3.3. Analogously to Remark 2.10 the relaxation Φ defined in (3.3) by the maximal monotone extension $\tilde{\Phi}$ as in (3.4) can be obtained by replacing h of (3.2) by means of its bi-conjugate, which is given by

$$h^{**}(r) = \begin{cases} \frac{\beta}{b} r & \text{for } r \in [0, b], \\ \frac{\beta}{a} r & \text{for } r \in [a, 0), \\ \infty & \text{for } r \in (-\infty, a) \cup (b, \infty). \end{cases}$$

Note that h^{**} is convex and a short computation shows that $(\partial h^{**})^{-1} = -\tilde{\Phi}$.

4. L^0: General case with regularization. In section 2 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. Here and in the following section Ω denotes a bounded domain with Lipschitz continuous boundary. We also employ 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

$$\min \quad j(u) + \frac{\alpha}{2}\, |\nabla u|^2_{L^2} + \beta \int_\Omega |u|_0^\varepsilon \, d\omega \text{ over } u \in X = H^1(\Omega), \tag{4.1}$$

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

$$u \to j(u) + \frac{\alpha}{2} |\nabla u|^2 \quad \text{is radially unbounded,} \tag{4.2}$$

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

In Remark 4.4 below we discuss how (2.27) relates to (4.2).

The sets of directional derivatives of $|\cdot|_0^\varepsilon$ is 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 4.1. *Problem* (4.1) *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 (4.2). We recall that $H^1(\Omega)$ is compactly embedded in $L^2(\Omega)$. Together with Lebesgue's bounded convergence theorem the existence proof follows with standard arguments. Computing directional derivatives, the necessary optimality condition follows. □

THEOREM 4.2. *Every weak subsequential limit* $\bar u \in X$ *of* $\{u_\varepsilon\}$ *as* $\varepsilon \to 0^+$ *is a solution to*

$$\min \quad j(u) + \frac{\alpha}{2}\, |\nabla u|^2_{L^2(\Omega)} + \beta \int_\Omega |u|_0 \, d\omega \;\; \text{over } u \in X = H^1(\Omega). \tag{4.3}$$

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*

$$\begin{cases} j'(\bar u) + \bar\lambda - \alpha\Delta\bar u = 0 \text{ in } X^*, \\ \lambda^\epsilon(\omega) \to \bar\lambda = 0 \text{ a.e. on } \{|\bar u| > 0\}, \\ \langle \bar\lambda, \phi\, \bar u\rangle_{X^*,X} \geq 0 \text{ for all } \phi \in C^1(\omega) \text{ with } \phi \geq 0. \end{cases} \tag{4.4}$$

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

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

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

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

and thus $\bar u$ is a solution to (4.3).

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 (4.4) follows. Since $u_\varepsilon \to \bar u$ we have the second equation in (4.4). Finally, taking the limit in $\langle \bar\lambda_\varepsilon, \phi\, u_\varepsilon\rangle_{X^*,X} \ge 0$ we obtain the last statement of (4.4). ∎

Remark 4.3. From Tchebycheff's inequality we have

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

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

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

Remark 4.4. The results of this section as well as of the following are directly applicable to optimal control problems of the form (2.27). In fact, in this case

$$j(u) = \frac{1}{2}|x - z|^2_{L^2(\Omega)} + \frac{\tilde\alpha}{2}|u|^2_{L^2(\Omega)},$$

where $Ex + Bu = g$, with E, B, g, z as in (2.27), and $\tilde\alpha > 0$. Then

$$j(u) = \frac{1}{2}|E^{-1}(g - Bu) - z|^2_{L^2(\Omega)},$$

and $j'(u) = B^*p$, where $E^*p = -(x - z)$.

5. L^p: Local regularization and monotone scheme. In this section we turn to minimization problems involving N_p-functionals with $p \in (0,1)$, where

$$N_p(f) = \int_\Omega |f(x)|^p\, dx.$$

We note that by Lebesgue's monotone convergence theorem one can argue that

$$\lim_{p\to 0^+} N_p(f) = N_0(f).$$

Throughout we restrict ourselves from the beginning to a class of problems for which we shall prove the convergence of a monotone scheme, and we consider

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

where $j : H^1(\Omega) \to \mathbb{R}^+$ is weakly lower semicontinuous, $\alpha > 0, \beta > 0$, and $\mathcal{C} \subset L^2(\Omega)$ is closed but not necessarily bounded.

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

$$\psi_\varepsilon(t) = \begin{cases} \frac{p}{2}\,\frac{t}{\varepsilon^{2-p}} + (1-\frac{p}{2})\varepsilon^p & \text{for } t \in [0,\varepsilon^2], \\ t^{\frac{p}{2}} & \text{for } t \in (\varepsilon^2,\infty), \end{cases} \tag{5.1}$$

where $\varepsilon > 0$, and we 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

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

where

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

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

$$N_p(u) \le N_{p,\varepsilon}(u) \le N_p(u) + \varepsilon^p|\Omega|, \tag{5.2}$$

where for the first inequality we used that $t^p \le \psi_\varepsilon(t^2)$ on $[0,\infty)$. The following lemma will be useful.

Lemma 5.1. *If $u_n \to u$ in $L^1(\Omega)$, then*

$$N_p(u_n) \to N_p(u) \;\; \text{and} \;\; N_{p,\varepsilon}(u_n) \to N_{p,\varepsilon}(u)$$

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 Hölder's inequality

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

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

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

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

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

Moreover, using Lemma 5.1 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

$$u_\varepsilon \to u \in L^1(\Omega) \quad \text{implies that } N_{p,\varepsilon}(u_\varepsilon) \to N_p(u). \tag{5.3}$$

Proposition 5.2. *For any $p \in (0, 1)$, problem $(P_{s,\varepsilon})$ admits a solution u_ε.*

Proof. For completeness we provide a 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. Using (5.2) we have

$$N_p(u_k^0) = N_p(u_k - u_k^1) \leq N_p(u_k) + N_p(u_k^1) \leq N_{p,\varepsilon}(u_k) + N_{p,\varepsilon}(u_k^1), \tag{5.4}$$

and hence $\{N_p(u_k^0)\}$ 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 5.1 and using weak lower semicontinuity of norms, it follows that u_ε is a solution. ∎

Proposition 5.3. *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 (5.3) and weak lower semicontinuity 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

$$j(u) = \frac{1}{2}|Ku - f|_Y^2, \tag{5.5}$$

where Y is a Hilbert space, $f \in Y$, and $K \in \mathcal{L}(H^1(\Omega), Y)$. Referring back to (2.27) the choice $K = -E^{-1}B$, $f = z - E^{-1}g$, and $Y = L^2(\Omega)$ can be considered as a special case in (5.5).

In the case that no constraints are present, i.e., $\mathcal{C} = L^2(\Omega)$, the necessary optimality condition for $(P_{s,\varepsilon})$ is given by

$$\begin{cases} -\alpha\Delta u + K^*Ku + \dfrac{\beta p}{\max(\varepsilon^{2-p}, |u|^{2-p})} \, u = K^*f & \text{in } \Omega, \\ \dfrac{\partial u}{\partial n} = 0 & \text{on } \partial\Omega, \end{cases} \tag{5.6}$$

where $|u| = |u(x)|$. To solve (5.6) we can use an iterative scheme and, given u_k, determine u_{k+1} from

$$-\alpha\Delta u_{k+1} + K^*Ku_{k+1} + \frac{\beta p}{\max(\varepsilon^{2-p}, |u_k|^{2-p})}\, u_{k+1} = K^*f. \tag{5.7}$$

By the Lax–Milgram theorem (5.7) admits a unique solution.

THEOREM 5.4. *The sequence $\{\mathcal{J}_\varepsilon(u_k)\}$ generated by* (5.7) *is strictly decreasing and*

$$\begin{aligned} &\frac{1}{2}\sum_{k=0}^{\infty}\left(\alpha|\nabla(u_{k+1}-u_k)|^2_{L^2} + |K(u_{k+1}-u_k)|^2_Y\right) \\ &\qquad + \frac{\beta p}{2}\sum_{k=0}^{\infty}\int_\Omega \frac{1}{\max(\varepsilon^{2-p}, |u_k|^{2-p})}|u_{k+1}-u_k|^2\,dx \le J_\varepsilon(u_0). \end{aligned} \tag{5.8}$$

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 (5.6).

Proof. Taking the inner product of (5.7) with $u_{k+1}-u_k$, for $k = 0, 1, \ldots$, we have

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

and hence

$$\begin{aligned} &\alpha\big(\nabla u_{k+1}, \nabla(u_{k+1}-u_k)\big)_{L^2(\Omega)} + \big(Ku_{k+1}-f, K(u_{k+1}-u_k)\big)_Y \\ &\quad + \left(\frac{\beta p}{\max(\varepsilon^{2-p}, |u_k|^{2-p})}u_{k+1}, u_{k+1}-u_k\right)_{L^2(\Omega)} = 0. \end{aligned}$$

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

$$\begin{aligned} &\alpha|\nabla u_{k+1}|^2_{L^2(\Omega)} - \alpha|\nabla u_k|^2_{L^2} + \alpha|\nabla u_{k+1}-\nabla u_k|^2_{L^2(\Omega)} + |Ku_{k+1}-f|^2_Y - |Ku_k-f|^2_Y \\ &\quad + |K(u_{k+1}-u_k)|^2_Y + \beta p\int_\Omega \frac{1}{\max(\varepsilon^{2-p}, |u_k|^{2-p})}|u_{k+1}-u_k|^2\,dx + G = 0, \end{aligned}$$

where

$$\begin{aligned} G &= \beta p\int_\Omega \frac{1}{\max(\varepsilon^{2-p}, |u_k|^{2-p})}(|u_{k+1}|^2 - |u_k|^2)\,dx \\ &= 2\beta\int_\Omega \psi'_\varepsilon(|u_k|^2)(|u_{k+1}|^2-|u_k|^2)\,dx \ge 2\beta\int_\Omega \big(\psi_\varepsilon(|u_{k+1}|^2) - \psi_\varepsilon(|u_k|^2)\big)\,dx. \end{aligned}$$

Combining these estimates we arrive at

$$\begin{aligned} &\alpha|\nabla u_{k+1}|^2_{L^2} + \alpha|\nabla(u_{k+1}-u_k)|^2_{L^2} + |Ku_{k+1}-f|^2_Y + |K(u_{k+1}-u_k)|^2_Y \\ &\quad + 2\beta\, N_{p,\varepsilon}(u_{k+1}) + \beta p\int_\Omega \frac{1}{\max(\varepsilon^{2-p}, |u_k|^{2-p})}|u_{k+1}-u_k|^2\,dx \\ &\quad \le \alpha|\nabla u_k|^2_{L^2[\Omega)} + |Ku_k-f|^2_Y + 2\beta\, N_{p,\varepsilon}(u_{k+1}), \end{aligned}$$

and thus

$$
\begin{aligned}
&\mathcal{J}_\varepsilon(u_{k+1}) + \frac{\alpha}{2}|\nabla(u_{k+1}-u_k)|^2 + \frac{1}{2}|K(u_{k+1}-u_k)|_Y^2 \\
&\qquad + \frac{\beta p}{2}\int_\Omega \frac{1}{\max(\varepsilon^{2-p}, |u_k|^{2-p})}|u_{k+1}-u_k|^2\,dx \le \mathcal{J}_\varepsilon(u_k).
\end{aligned}
\tag{5.9}
$$

This implies that for every k

$$
\begin{aligned}
&\mathcal{J}_\varepsilon(u_{k+1}) + \frac{1}{2}\sum_{i=0}^{k}\left(\alpha|\nabla(u_{i+1}-u_i)|_{L^2}^2 + |K(u_{i+1}-u_i)|_Y^2\right) \\
&\qquad + \frac{\beta p}{2}\sum_{i=0}^{k}\int_\Omega \frac{1}{\max(\varepsilon^{2-p}, |u_i|^{2-p})}|u_{i+1}-u_i|^2\,dx \le \mathcal{J}_\varepsilon(u_0)
\end{aligned}
\tag{5.10}
$$

and estimate (5.8) follows. From (5.9) 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 (5.6).

From (5.10) we deduce that $\{|\nabla u_k|_{L^2}\}$ and $\{N_{p,\varepsilon}(|u_k|)\}$ are bounded sequences. Since $N_p(u_k) \le N_{p,\varepsilon}(u_k)$ by (5.2), it follows that $\{N_p(u_k)\}$ is bounded as well. Decomposing $u_k = u_k^0 + u_k^1$ as in the proof of Proposition 5.2, we find that the constant parts $\{u_k^0\}$ are bounded (compare (5.4)), 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 (5.8) 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 (5.7) we find that u_ε is a solution to (5.6). ◻

Acknowledgment. The authors acknowledge many useful comments by the anonymous referees.

REFERENCES

[1] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, Springer, New York, 2011.

[2] F. Bonnans and E. Casas, *An extension of Pontryagin's principle for state-constrained optimal control of semilinear elliptic equations and variational inequalities*, SIAM J. Control Optim., 33 (1995), pp. 274–298.

[3] K. Bredies, D. A. Lorenz, and S. Reiterer, *Minimization of Non-smooth, Non-convex Functionals by Iterative Thresholding*, preprint, 2009.

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

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

[6] E. Casas, R. Herzog, and G. Wachsmuth, *Optimality conditions and error analysis of semilinear elliptic control problems with l1 cost functional*, SIAM J. Optim., 22 (2012), pp. 795–820.

[7] E. Casas and J. Yong, *Maximum principle for state-constrained optimal control problems governed by quasilinear elliptic equations*, Differential Integral Equations, 8 (1995), pp. 1–19.

[8] C. Clason and K. Kunisch, *A duality-based approach to elliptic control problems in non-reflexive Banach spaces*, ESAIM Control Optim. Calc. Var., 17 (2011), pp. 243–266.

[9] I. Daubechies, M. Defrise, and C. D. Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math., 57 (2004), pp. 1413–1457.

[10] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.

[11] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, SIAM, Philadelphia, 1999.

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

[13] G. Stadler, *Elliptic optimal control problems with $L1$-control cost and applications for the placement of control devices*, Comput. Optim. Appl., 44 (2007), pp. 159–181.

[14] M. Hintermüller and T. Wu, *Nonconvex TV^q-Models in Image Restoration: Analysis and a Trust-Region Regularization Based Superlinearly Convergent Solver*, preprint, 2012.

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

[16] K. Ito and K. Kunisch, *A variational approach to sparsity optimization based on Lagrange multiplier theory*, Inverse Problems, 30 (2014), DOI:10.1088/0266-5611/30/1/015001.

[17] B. Jin, T. Khan, and P. Maass, *A reconstruction algorithm for electrical impedance tomography based on sparsity regularization*, Int. J. Numer. Math. Engrg., 89 (2012), pp. 337–353.

[18] E. J. McShane, *Relaxed controls and variational problems*, SIAM J. Control, 5 (1967), pp. 438–485.

[19] M. Nikolova, *Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers*, SIAM J. Numer. Anal., 40 (2002), pp. 965–994.

[20] R. Schowalter, *Monotone Operators in Banach Spaces and Nonlinear Partial Differential Equations*, American Mathematical Society, Providence, RI, 1997.

[21] J. Warga, *Relaxed variational problems*, J. Math. Anal. Appl., 4 (1962), pp. 111–128.

[22] C. A. Zarzer, *On Tikhonov regularization with non-convex sparsity constraints*, Inverse Problems, 25 (2009), 25006.