Optimal control problems with $L^0(\Omega)$ constraints: maximum principle and proximal gradient method

Daniel Wachsmuth[1]

Abstract

We investigate optimal control problems with L^0 constraints, which restrict the measure of the support of the controls. We prove necessary optimality conditions of Pontryagin maximum principle type. Here, a special control perturbation is used that respects the L^0 constraint. First, the maximum principle is obtained in integral form, which is then turned into a pointwise form. In addition, an optimization algorithm of proximal gradient type is analyzed. Under some assumptions, the sequence of iterates contains strongly converging subsequences, whose limits are feasible and satisfy a subset of the necessary optimality conditions.

Keywords Sparse optimal control · L^0 constraints · Pontryagin maximum principle · Proximal gradient method

Mathematics Subject Classification 49M20 · 49K20

1 Introduction

We are interested in the following optimal control problem written as an optimation problem:

$$\min_{u \in L^2(\Omega)} f(u) + \frac{\alpha}{2} \|u\|_{L^2(\Omega)}^2 \tag{1.1}$$

subject to

$$\|u\|_0 \le \tau. \tag{1.2}$$

✉ Daniel Wachsmuth
  daniel.wachsmuth@mathematik.uni-wuerzburg.de

1   Institut für Mathematik, Universität Würzburg, 97074 Würzburg, Germany

Here, $\Omega \subseteq \mathbb{R}^d$ is an open set supplied with the Lebesgue measure, $f : L^2(\Omega) \to \mathbb{R}$ abstracts the state equation and smooth ingredients of the control problem, $\alpha \geq 0$ is a parameter. The constraint (1.2) uses the so-called L^0 norm (which is—of course— not a norm) that is defined for measurable $u : \Omega \to \mathbb{R}$ by

$$\|u\|_0 := \operatorname{meas}\{x : u(x) \neq 0\}.$$

Of course, $\tau \in (0, \operatorname{meas}(\Omega))$ is a meaningful restriction.

The motivation to study such problems comes from sparse control: Find a control with small support, in our case: with prescribed size of support. The main challenge is the discontinuity and non-convexity of the $\| \cdot \|_0$-functional: Methods from differentiable or convex optimization are not applicable. In addition, due to the lack of weak lower continuity it is not possible to ensure existence of solutions in spaces of integrable functions. Nevertheless, we can look into optimality conditions that need to be satisfied at a solution. In order to study necessary optimality conditions, we will employ the Pontryagin maximum principle, which is first obtained in integral form, and then turned into a pointwise condition by means of natural arguments.

Let us mention related works. Optimal control problems with L^0 norm of the control in the cost function were investigated in [10, 15]. An actuator design problem is studied in [11]: the controlled source term in the equation is $\chi_\omega u$, where χ_ω is the characteristic function of ω, and the subset ω and the control u are optimization variables. An additional volume constraint is posed on ω, which is equivalent to a L^0 constraint on $\chi_\omega u$. In that work, shape calculus and topological derivatives with respect to ω are studied. Unfortunately, no optimality conditions involving these topological derivatives are given, which could be compared to our results. This is subject to future work. In the recent work [4], a shape optimization problem is turned into a problem with L^0 constraints. There the control problem is posed in $W^{1,p}$, and offers different challenges than the setting considered here. That work will become relevant if one wants to study the regularization of (1.1)–(1.2) in $W^{1,p}$ spaces, which would guarantee existence of solutions due to the compact embedding of $W^{1,p}$ in L^p.

In this article, we will prove optimality conditions of Pontryagin maximum principle type. Related works can be found, e.g., in [5, 6, 14]. Those results are not directly applicable in our situation, since they do not cover L^0 constraints. We will use a modification of the control perturbations considered in [5, 6, 14] that is adapted to the L^0 constraints. These will give the maximum principle in integral form, see Theorem 4.4. In order to turn it into pointwise conditions in Theorem 4.5, we study integral minimization problems in Sect. 3.

In Sect. 5, we investigate an proximal gradient type algorithm, which extends our earlier works [13, 15], where optimization problems with L^0 and L^p, $p \in (0, 1)$, functionals were considered. Due to the simple nature of its sub-problems, this method is easy to implement. Other methods in finite-dimensional L^0 constrained (or cardinality constrained) optimization include augmented Lagrangian methods [12] and DC-based reformulations [8]. We will prove some convergence results for the proximal gradient method. As it turns out, limit points of

iterates do not satisfy the necessary condition Theorem 5.2 but only a subset of those, see Theorem 5.7. We hope that this work initiates further research on algorithms with L^0 constraints in an infinite-dimensional setting.

Notation We will frequently use the following notation: For a measurable set A, we denote its characteristic function by χ_A. The integrand in the L^0 norm is abbreviated by

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

Then $\|u\|_0 = \int_\Omega |u(x)|_0 \, dx$. Note that $u \mapsto |u|_0$ is neither continuous nor convex but lower semicontinuous. In addition, $u \mapsto \|u\|_0$ as mapping from $L^p(\Omega)$ to $\mathbb{R}$ is lower semicontintinuous but not weakly lower semicontinuous. Moreover, we will denote the support of the measurable function u by

$$\text{supp} u := \{x \in \Omega : u(x) \neq 0\}.$$

2 Maximum principle for control of ordinary differential equations

Let us briefly and formally derive the maximum principle for an optimal control problem subject to ordinary differential equations with constraint $\|u\|_0 \leq \tau$, which serves as benchmark for more general situations. For illustration, let us consider the following control problem in Mayer form: Minimize

$$l(x(T))$$

subject to

$$\begin{aligned} &x'(t) = f(t, x(t), u(t)) \text{ a.e. on } (0, T), \\ &x(0) = x_0, \\ &u(t) \in U \text{ for almost all } t \in (0, T), \end{aligned}$$

and

$$\|u\|_0 \leq \tau.$$

Here, $T > 0$ is fixed, and $x : (0, T) \to \mathbb{R}^n$ and $u : (0, T) \to \mathbb{R}$ are the state and control. The functions $f : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}$ and $l : \mathbb{R}^n \to \mathbb{R}$ are assumed to be smooth for simplicity. Employing a standard procedure, the constraint $\|u\|_0 \leq \tau$ can be written equivalently as an additional end-point constraint on the artificial state x_{n+1} as follows:

$$x_{n+1}(0) = 0, \ x'_{n+1}(t) = |u(t)|_0 \text{ a.e. on } (0, T), \ x_{n+1}(T) \leq \tau.$$

Let us set $\tilde{f}(t, x, u) := (f(t, x, u), |u|_0)$. Then the classical maximum principle for an optimal control $\bar{u}$ with state $(\bar{x}, \bar{x}_{n+1})$ and adjoint $(\bar{p}, \bar{p}_{n+1}) \in \mathbb{R}^{n+1}$ is: there are $(\lambda_0, \lambda_{n+1}) \neq 0$, $\lambda_0 \geq 0$, such that the following conditions are satisfied:

$$\bar{u}(t) = \arg\max_{u \in U} H(t, \bar{x}(t), u, \bar{p}(t)) + \bar{p}_{n+1}(t)|u|_0,$$

where $H(t, x, u, p) := p^T f(t, x, u)$ is the Hamiltonian of the original problem, and $\bar{p}$ solves the adjoint system

$$-\bar{p}(T) = \lambda_0 l'(\bar{x}(T)), \quad -\bar{p}'(t) = f_x(t, \bar{x}(t), \bar{u}(t))^T \bar{p}(t) \text{ a.e. on } (0, T)$$

and

$$-\bar{p}_{n+1}(T) = \lambda_{n+1}, \quad -\bar{p}'_{n+1}(t) = 0,$$
$$\lambda_{n+1} \geq 0, \quad \lambda_{n+1}(\bar{x}_{n+1}(T) - \tau) = 0.$$

Hence, $\bar{p}_{n+1}$ is constant, $\bar{p}_{n+1} \leq 0$, and $-\bar{p}_{n+1}$ can be interpreted as Lagrange multiplier to the constraint $\|u\|_0 \leq \tau$. For a precise formulation of the maximum principle, we refer to [9]. In order to obtain the system in qualified form, i.e., $\lambda_0 > 0$, one needs additional conditions (constraint qualifications).

Summarizing the above considerations, the following two conditions will serve as necessary optimality conditions: $\bar{u}(t)$ maximizes the penalized Hamiltonian, i.e.,

$$\bar{u}(t) = \arg\max_{u \in U} H(t, \bar{x}(t), u, \bar{p}(t)) - \lambda|u|_0, \tag{2.1}$$

and $\lambda \geq 0$ satisfies the complementarity condition

$$\lambda(\|\bar{u}\|_0 - \tau) = 0. \tag{2.2}$$

Let us now transfer these results to optimal control problems, where the control is no longer defined on a subset of the real line but defined on a set $\Omega \subseteq \mathbb{R}^d$, $d > 1$. For illustration, let now $\Omega \subseteq \mathbb{R}^d$, $d > 1$, be a bounded domain. As above, we want to translate the control constraint $\|u\|_0 \leq \tau$ to an auxiliary state constraint. In fact, if we define y_0 as the weak solution in $H^1(\Omega)$ of the auxiliary state equation

$$-\Delta y + y = |u|_0, \quad \frac{\partial y}{\partial n} = 0$$

then it follows $\int_\Omega y_0 \, dx = \int_\Omega |u|_0 \, dx$, and the control constraint $\|u\|_0 \leq \tau$ is equivalent to the constraint $\int_\Omega y_0 \, dx \leq \tau$ on the auxiliary state y. In [6], the maximum principle for problems with elliptic partial differential equations was obtained. In order to get a system in qualified form ($\lambda_0 > 0$) strong stability is used: the optimal value function has to be locally Lipschitz continuous with respect to the parameter τ. To the best of our knowledge, such a result is not available in the literature for the L^0 constraints considered here.

Thus, we will proceed differently. We will not formulate the integral control constraint as a state constraint. Rather we will modify the technique of [6] to only consider perturbations that satisfy the constraint. In this way, we get a maximum principle in integral form satisfied for all functions v with $\|v\|_0 \leq \tau$. This integral maximum principle can be translated into a pointwise one. In that way, we get the final system in qualified form while circumventing the strong stability requirement.

3 Optimality conditions for integral functionals

First, we are going to derive optimality conditions for integral functionals. This is later used to transform the maximum principle from integral to pointwise form. We will consider integral functionals generated by normal integrands. In this section, let $\Omega \subseteq \mathbb{R}^d$ be a Lebesgue measurable set.

Definition 3.1 The function $f : \Omega \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is called *normal integrand* if there exist Caratheodory functions $(f_n)_{n \in \mathbb{N}}$ such that for all u and almost all $x \in \Omega$

$$f(x, u) = \sup_n f_n(x, u).$$

for all u and almost all $x \in \Omega$.

This definition is from [2, Def. 1] with equivalent characterizations in [2, Thm. 2]. We will now prove optimality conditions for the following problem: Minimize

$$\int_\Omega g(x, u(x)) \, dx \qquad (3.1a)$$

subject to the constraint

$$u \in \mathcal{U}_\tau := \{u \text{ measurable} : \ \|u\|_0 \le \tau\}. \qquad (3.1b)$$

Here, the minimization is over all measurable u such that $g(\cdot, u)$ is integrable. Clearly, if $\bar{u}$ is a solution of (3.1) and $\bar{u}(x) \ne 0$ then $g(x, \bar{u}(x)) = \inf_{v \in \mathbb{R}} g(x, v)$. The latter function will play an important role in the subsequent analysis. We work with the following assumption.

Assumption 1

1. $g : \Omega \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is a normal integrand,
2. $g(\cdot, 0)$ is integrable,
3. $\tau \in (0, \text{meas}(\Omega))$.

Let us define the non-positive function $\tilde{v} : \Omega \to \mathbb{R} \cup \{-\infty\}$ by

$$\tilde{v}(x) := \inf_{v \in \mathbb{R}} g(x, v) - g(x, 0). \qquad (3.2)$$

We start with a technical lemma that helps to prove integrability of $\tilde{v}$ under suitable assumptions.

Lemma 3.2 *Let* $g : \Omega \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ *be a normal integrand. Then there are measurable functions* v_n *and* u_n *such that*

$$v_n(x) = \inf_{|u| \le n} g(x, u) - g(x, 0) = g(x, u_n(x)) - g(x, 0), \quad |u_n(x)| \le n$$

for almost all $x \in \Omega$. In addition, $\tilde{v}$ is measurable.

Proof Let (g_n) be Caratheodory functions such that $g(x, u) = \sup_n g_n(x, u)$ for all u and almost all $x \in \Omega$. Let us define the set-valued mapping

$$E(x) := \{(v, t) : \ g(x, v) - g(x, 0) \le t\},$$

so $E(x)$ is the epi-graph of $v \mapsto g(x, v) - g(x, 0)$. Then it holds

$$E(x) = \bigcap_n \{(v, t) : \ g_n(x, v) - t \le g(x, 0)\}.$$

Each of the set-valued mappings in the intersection is measurable by [1, Thm. 8.2.9], so E is measurable by [1, Thm. 8.2.4]. Using [1, Thm. 8.2.11], we get the measurability of u_n and v_n. Measurability of $\tilde{v}$ is a consequence of $\tilde{v}(x) = \inf_n v_n(x)$. $\qquad \square$

Using the function $\tilde{v}$ from (3.2), we define the sets

$$\Omega_{<s} := \{x : \ \tilde{v}(x) < s\}, \ \Omega_{\le s} := \{x : \ \tilde{v}(x) \le s\}. \tag{3.3}$$

Lemma 3.3 *Let $u \in \mathcal{U}_\tau$ be given such that $g(\cdot, u)$ is integrable. Let $s \le 0$ and $S \subseteq \Omega$ with $\mathrm{meas}(S) = \tau$ be such that $\tilde{v}$ (see (3.2)) is integrable on S and*

$$\Omega_{<s} \subseteq S \subseteq \Omega_{\le s}.$$

Then it holds

$$\int_\Omega g(x, u(x)) - g(x, 0) \, \mathrm{d}x \ge \int_S \tilde{v}(x) \, \mathrm{d}x.$$

This inequality is satisfied with equality only if the following conditions are satisfied:

1. $g(x, u(x)) - g(x, 0) = \tilde{v}(x)$ *for almost all $x \in \mathrm{supp}u$,*
2. $\Omega_{<s} \subseteq \mathrm{supp}u \subseteq \Omega_{\le s}$,
3. $s = 0$ *or* $\mathrm{meas}(\mathrm{supp}u) = \tau$.

Proof Let $A := \mathrm{supp}u$. Then $\mathrm{meas}(A) \le \tau = \mathrm{meas}(S)$, and it follows $\mathrm{meas}(A \setminus S) \le \mathrm{meas}(S \setminus A)$. Using (3.3), we estimate

$$\int_\Omega g(x, u(x)) - g(x, 0) \, dx \geq \int_A \tilde{v}(x) \, dx$$
$$= \int_{A \cap S} \tilde{v}(x) \, dx + \int_{A \setminus S} \tilde{v}(x) \, dx$$
$$\geq \int_{A \cap S} \tilde{v}(x) \, dx + s \, \mathrm{meas}(A \setminus S)$$
$$\geq \int_{A \cap S} \tilde{v}(x) \, dx + s \, \mathrm{meas}(S \setminus A)$$
$$\geq \int_S \tilde{v}(x) \, dx.$$

Equality in the above chain of inequalities is obtained only if (a) $g(\cdot, u) - g(\cdot, 0) = \tilde{v}$ on A, (b) $\tilde{v} = s$ on $A \setminus S$, hence $A \subseteq \Omega_{\leq s}$, (c) $s(\mathrm{meas}(A \setminus S) - \mathrm{meas}(S \setminus A)) = 0$, and (d) $s \, \mathrm{meas}(S \setminus A) = \int_{S \setminus A} \tilde{v} \, dx$. Condition (d) implies $\tilde{v} = s$ on $S \setminus A$, hence $\Omega_{<s} \subseteq A$. If $s \neq 0$ then condition (c) implies $\mathrm{meas}(A) = \tau$. □

With the help of these sets, we can fully characterize the solutions of (3.1).

Theorem 3.4 *Let Assumption* 1 *be satisfied. Then $\bar{u}$ is a solution of* (3.1) *if and only if there are $s \leq 0$ and $A \subseteq \Omega$ with* $\mathrm{meas}(A) = \tau$ *such that*

$$\Omega_{<s} \subseteq \mathrm{supp}\bar{u} \subseteq A \subseteq \Omega_{\leq s}, \tag{3.4}$$

$\tilde{v}$ is integrable on A, and

$$g(x, \bar{u}(x)) - g(x, 0) = \tilde{v}(x) \text{ for almost all } x \in A, \tag{3.5}$$

where $\tilde{v}$ is defined in (3.2).

Proof Let $\bar{u}$ be a solution of (3.1). Let (v_n) and (u_n) be given by Lemma 3.2. By construction, $(v_n(x))$ is monotonically decreasing and $v_n(x) \to \tilde{v}(x)$ for almost all $x \in \Omega$. Let $B \subseteq \Omega$ with $\mathrm{meas}(B) \leq \tau$. We want to show that $\chi_B u_n$ is feasible for (3.1). It remains to argue that $g(\cdot, \chi_B u_n)$ is integrable. If the negative part of $g(\cdot, \chi_B u_n)$ would not be integrable, then problem (3.1) would be unsolvable, as we could find subsets $B_k \subseteq B$ such that $\int_\Omega g(x, \chi_{B_k} u_n) \, dx \to -\infty$ for $k \to \infty$. So the negative part of $g(\cdot, \chi_B u_n)$ is integrable, and the integrability of $g(\cdot, \chi_B u_n)$ is a consequence of $g(x, \chi_B u_n(x)) - g(x, 0) \leq 0$ for almost all x.

Then $\chi_B u_n$ is feasible for (3.1), which implies

$$0 \geq \int_B v_n \, dx = \int_\Omega g(x, \chi_B u_n(x)) - g(x, 0) \, dx \geq \int_\Omega g(x, \bar{u}(x)) - g(x, 0) \, dx.$$

By the monotone convergence theorem, it follows that $\tilde{v}$ is integrable on B and

$$\int_B \tilde{v} \, dx \geq \int_\Omega g(x, \bar{u}(x)) - g(x, 0) \, dx. \tag{3.6}$$

The increasing functions $s \mapsto \text{meas}(\Omega_{<s})$ and $s \mapsto \text{meas}(\Omega_{\leq s})$ are continuous from the left and from the right, respectively. Given τ, there is a uniquely determined $s \leq 0$ such that $\text{meas}(\Omega_{<s}) \leq \tau \leq \text{meas}(\Omega_{\leq s})$. Since the measure space is non-atomic, the celebrated Sierpiński theorem implies that there is $S \subseteq \Omega$ such that $\Omega_{<s} \subseteq S \subseteq \Omega_{\leq s}$ and $\text{meas}(S) = \tau$.

By the first part of the proof, $\tilde{v}$ is integrable on S. Then s and S satisfy the requirements of Lemma 3.3. Using Lemma 3.3 and (3.6), we get

$$\int_\Omega g(x, \bar{u}(x)) - g(x, 0) \, dx \geq \int_S \tilde{v}(x) \, dx \geq \int_\Omega g(x, \bar{u}(x)) - g(x, 0) \, dx.$$

Hence, the inequality of Lemma 3.3 is satisfied with equality, which implies $\Omega_{<s} \subseteq \text{supp}\,\bar{u} \subseteq \Omega_{\leq s}$. It remains to utilize that $s = 0$ or $\text{meas}(\text{supp}\,\bar{u}) = \tau$. If $\text{meas}(\bar{u}) = \tau$ then (3.4) and (3.5) are satisfied with $A := \text{supp}\,\bar{u}$. If $\text{meas}(\text{supp}\,\bar{u}) < \tau$ then $s = 0$, and we can find a set A with $\text{meas}(A) = \tau$ and $\text{supp}\,\bar{u} \subseteq A \subseteq \Omega = \Omega_{\leq 0}$, which is (3.4). Using Lemma 3.3 and $\bar{u}(x) = 0$ on $A \setminus \text{supp}\,\bar{u}$, we see that (3.5) is satisfied.

Let now $\bar{u}, s, A$ satisfy (3.4) and (3.5) such that $\tilde{v}$ is integrable on A. Let $u \in \mathcal{U}_\tau$. Then by Lemma 3.3 with $S := A$ we find

$$\int_\Omega g(x, u(x)) - g(x, 0) \, dx \geq \int_A \tilde{v}(x) \, dx = \int_A g(x, \bar{u}(x)) - g(x, 0) \, dx$$
$$= \int_\Omega g(x, \bar{u}(x)) - g(x, 0) \, dx,$$

and $\bar{u}$ solves (3.1). □

Corollary 3.5 *Let Assumption 1 be satisfied. Let $\bar{u}$ be a solution of (3.1). Let $s \leq 0$ be given by Theorem 3.4. Then for almost all $x \in \Omega$*

$$|\bar{u}(x)|_0 \cdot (\tilde{v}(x) - s) \leq 0.$$

Proof This follows from Theorem 3.4, (3.4): If $\bar{u}(x) \neq 0$, then $\tilde{v}(x) \leq s$. □

Let us define the value function of (3.1) by

$$V(\tau) := \inf_{u \in \mathcal{U}_\tau} \int_\Omega g(x, u(x)) \, dx.$$

Using the above characterization of solutions, we have the following strong stability result.

Lemma 3.6 *Let Assumption 1 be satisfied. Let $\tau, \tau' \in (0, \text{meas}(\Omega))$ with $\tau < \tau'$ be given. Then $0 \leq V(\tau) - V(\tau') \leq |s|(\tau - \tau')$, where s is associated to τ by Theorem 3.4.*

Proof Let $u_\tau, u_{\tau'}$ be solutions to τ, τ'. Due to Theorem 3.4 there are s, s', A, A' such that $\text{meas}(A) = \tau$, $\text{meas}(A') = \tau'$, and

$$\Omega_{<s} \subseteq A \subseteq \Omega_{\leq s}, \quad \Omega_{<s'} \subseteq A' \subseteq \Omega_{\leq s}.$$

If $s < s'$ then $\Omega_{\leq s} \subseteq \Omega_{<s'}$ and $A \subseteq A'$, which implies

$$
\begin{aligned}
\int_\Omega g(x, u_\tau(x))\,\mathrm{d}x - \int_\Omega g(x, u_{\tau'}(x))\,\mathrm{d}x &= \int_A \tilde{v}\,\mathrm{d}x - \int_{A'} \tilde{v}\,\mathrm{d}x \\
&= -\int_{A'\setminus A} \tilde{v}\,\mathrm{d}x \\
&\leq -s\,\mathrm{meas}(A' \setminus A) = -s(\tau' - \tau).
\end{aligned}
$$

If $s = s'$ then

$$
\begin{aligned}
&\int_\Omega g(x, u_\tau(x))\,\mathrm{d}x - \int_\Omega g(x, u_{\tau'}(x))\,\mathrm{d}x \\
&= \int_{A\setminus\Omega_{<s}} \tilde{v}\,\mathrm{d}x - \int_{A'\setminus\Omega_{<s}} \tilde{v}\,\mathrm{d}x \\
&= s(\mathrm{meas}(A \setminus \Omega_{<s}) - \mathrm{meas}(A' \setminus \Omega_{<s})) = -s(\tau' - \tau),
\end{aligned}
$$

resulting in the same estimate. $\qquad\square$

In addition, we obtain the following result, which says that $-s$ can be interpreted as Lagrange multiplier to the constraint $\|u\|_0 \leq \tau$.

Corollary 3.7 *Let Assumption 1 be satisfied. Let $\bar{u}$ be a solution of (3.1). Let $s \leq 0$ be given by Theorem 3.4. Then we have*

$$s \cdot (\tau - \|\bar{u}\|_0) = 0. \tag{3.7}$$

Proof Suppose $\|\bar{u}\|_0 < \tau$. By Theorem 3.4 there is A with $\mathrm{meas}(A) = \tau$ and $\mathrm{supp}\,\bar{u} \subseteq A \subseteq \Omega_{\leq s}$. Due to (3.5), $\tilde{v} = 0$ on $A \setminus \mathrm{supp}\,\bar{u} \subseteq \Omega_{\leq s}$, where $A \setminus \mathrm{supp}\,\bar{u}$ has positive measure. Hence, $s = 0$ follows by definition of $\Omega_{\leq s}$, see (3.3). $\qquad\square$

Furthermore, $\bar{u}$ is a solution of unconstrained penalized problems, where $-s$ plays the role of a penalization parameter.

Corollary 3.8 *Let Assumption 1 be satisfied. Let $\bar{u}$ be a solution of (3.1). Let $\lambda := -s \geq 0$, where s is given by Theorem 3.4. Then $\bar{u}$ is a solution of*

$$\min_u \int_\Omega g(x, u(x)) + \lambda|u(x)|_0\,\mathrm{d}x.$$

and a solution of

$$\min_u \int_\Omega g(x, u(x))\,\mathrm{d}x + \lambda(\|u\|_0 - \tau)^+.$$

Proof Let s and A be as in Theorem 3.4. Let u be measurable and set $B := \operatorname{supp} u$. As in the proof of Lemma 3.3, we get

$$\int_\Omega g(x, u(x)) - g(x, 0)\, dx \geq \int_B \tilde{v}\, dx \geq \int_{A \cap B} \tilde{v}(x)\, dx + s\operatorname{meas}(B \setminus A)$$

and

$$\int_{A \cap B} \tilde{v}(x)\, dx + s\operatorname{meas}(A \setminus B) \geq \int_A \tilde{v}\, dx = \int_\Omega g(x, \bar{u}(x)) - g(x, 0)\, dx,$$

which results in

$$\int_\Omega g(x, u(x)) - g(x, \bar{u}(x)) \geq s(\operatorname{meas}(B \setminus A) - \operatorname{meas}(A \setminus B)).$$

We proceed with

$$
\begin{aligned}
s(\operatorname{meas}(B \setminus A) - \operatorname{meas}(A \setminus B)) &= s(\operatorname{meas}(B \setminus A) - \operatorname{meas}(A \setminus B) + \|\bar{u}\|_0 - \|\bar{u}\|_0) \\
&\geq s(\operatorname{meas}(B \setminus A) + \operatorname{meas}(A \cap B) - \|\bar{u}\|_0) \\
&= s(\|u\|_0 - \|\bar{u}\|_0),
\end{aligned}
$$

where we used $\|\bar{u}\|_0 \leq \operatorname{meas}(A)$, $\|u\|_0 = \operatorname{meas}(B)$, and $s \leq 0$. This proves the first claim. Using the result of Corollary 3.7 and $s \leq 0$, we get

$$s(\operatorname{meas}(B \setminus A) - \operatorname{meas}(A \setminus B)) \geq s(\|u\|_0 - \tau) \geq s(\|u\|_0 - \tau)^+,$$

which proves the second claim. $\qquad\square$

Let us prove the following converse result.

Corollary 3.9 *Let Assumption* 1 *be satisfied. Let* $\lambda' \geq 0$. *Let* $\bar{u}$ *with* $\|\bar{u}\|_0 = \tau$ *be a solution of*

$$\min_v \int_\Omega g(x, v(x)) + \lambda' |v(x)|_0\, dx.$$

Then $\bar{u}$ *solves* (3.1).

Proof Let u be given with $\|u\|_0 \leq \tau$. By optimality of $\bar{u}$, we have

$$
\begin{aligned}
\int_\Omega g(x, \bar{u}(x))\, dx + \lambda' \tau &= \int_\Omega g(x, \bar{u}(x)) + \lambda' |\bar{u}(x)|_0\, dx \\
&\leq \int_\Omega g(x, u(x)) + \lambda' |u(x)|_0\, dx \leq \int_\Omega g(x, u(x))\, dx + \lambda' \tau,
\end{aligned}
$$

which implies the claim. $\qquad\square$

Let us close the section with the following observation: Every minimum of the integral functional $\int_\Omega g(x, u(x))\, dx$ subject to the constraint $u \in \mathcal{U}_\tau \cap L^p(\Omega)$ is a solution of (3.1).

Theorem 3.10 *Let Assumption 1 be satisfied. Let $p \in [1, \infty]$. Let $\bar{u} \in L^p(\Omega)$ be a solution of*

$$\min_{u \in \mathcal{U}_\tau \cap L^p(\Omega)} \int_\Omega g(x, u(x))\, dx.$$

Then $\bar{u}$ solves (3.1).

Proof Let (u_n) and (v_n) be given by Lemma 3.2, which implies $u_n \in L^\infty(\Omega)$ for all n. Let $B \subseteq \Omega$ with meas$(B) \leq \tau$ be given, hence $\chi_B u_n \in L^p(\Omega)$ for all n. Arguing as in the proof of Theorem 3.4, we get $\int_B v_n\, dx \to \int_B \tilde{v}\, dx \geq \int_\Omega g(x, \bar{u}(x)) - g(x, 0)\, dx$ by monotone convergence, see (3.6). Let now u be feasible for (3.1). Let $B := \mathrm{supp}u$. Then

$$\int_\Omega g(x, u(x)) - g(x, 0)\, dx \geq \int_B \tilde{v}\, dx \geq \int_\Omega g(x, \bar{u}(x)) - g(x, 0)\, dx,$$

hence $\bar{u}$ solves (3.1) as well. □

Remark 3.11 All the results of this section are valid in the more general situation of a non-atomic, complete, σ-finite measure space.

4 Optimal control of elliptic partial differential equation with L^0 constraint

In this section, we consider the following optimal control problem: Minimize

$$\int_\Omega L(x, y_u(x), u(x))\, dx \tag{4.1a}$$

subject to

$$\|u\|_0 \leq \tau, \tag{4.1b}$$

where y_u is the weak solution of the equation

$$\begin{aligned} (Ay)(x) &= f(x, y(x), u(x)) \quad \text{on } \Omega \\ y &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{4.1c}$$

We impose the following assumption on the data of this problem:

Assumption 2

1. Ω is an open and bounded domain in $\mathbb{R}^d$, $d \in \{2, 3\}$, with Lipschitz boundary $\partial\Omega$. Let $\tau \in (0, \text{meas}(\Omega))$.

2. A denotes a second-order elliptic operator in Ω of the type

$$Ay = -\sum_{i,j=1}^{d} \partial_{x_j}(a_{ij}(x)\partial_{x_i}y)$$

   with coefficients $a_{ij} \in C(\bar{\Omega})$. In addition, there is $\Lambda > 0$ such that for almost all $x \in \Omega$

$$\sum_{i,j=1}^{d} a_{ij}(x)\xi_i\xi_j \geq \Lambda|\xi|^2 \quad \forall \xi \in \mathbb{R}^d.$$

3. The functions $f, L : \Omega \times \mathbb{R} \times \mathbb{R}$ are Caratheodory functions, i.e., $x \mapsto f(x, y, u)$ and $x \mapsto L(x, y, u)$ are measurable for all $y, u \in \mathbb{R}$, and $(u, y) \mapsto f(x, y, u)$ and $(u, y) \mapsto L(x, y, u)$ are continuous for almost all $x \in \Omega$. We assume that f, L are continuously differentiable with respect to y for almost all $x \in \Omega$ and all $u \in \mathbb{R}$ with $f_y(x, y, u) \leq 0$. In addition, for all $M > 0$ there are non-negative $a_M \in L^1(\Omega)$, $b_M \in \mathbb{R}$, $c_M \in L^2(\Omega)$ such that for almost all $x \in \Omega$

$$|L(x, y, u)| + |L_y(x, y, u)| \leq a_M(x) + b_M|u|^2 \quad \forall|y| \leq M$$

   and

$$|f(x, y, u)| + |f_y(x, y, u)| \leq c_M(x) + b_M|u| \quad \forall|y| \leq M,$$

   where f_y, L_y denote the partial derivatives of f, L with respect to y.

Let us briefly comment on those assumptions. The conditions on the differential equation are to ensure $W^{1,p}$ regularity of weak solutions y of (4.1c) for some $p > d$, which guarantees $y \in L^\infty(\Omega)$. The conditions on L and f ensure that the Nemyzki operators induced by them are continuous (and differentiable with respect to y) from $L^\infty(\Omega) \times L^2(\Omega)$ to $L^1(\Omega)$ and $L^2(\Omega)$, respectively. We opted for this set of conditions in order to be able to use the results of [6] on regularity of solutions of partial differential equations. This allows us to focus on the L^0 constraints. Of course, other settings are possible (e.g., control constraints, other types of boundary conditions, parabolic equations).

As consequence of the assumptions, we have the following solvability and regularity result for (4.1c).

Theorem 4.1 *Let Assumption 2 be satisfied. Let $u \in L^2(\Omega)$ be given. Then there is a uniquely determined $y_u \in W_0^{1,p}(\Omega)$ solving the Eq. (4.1c), where $p > d$.*

Proof This is a consequence of [6, Theorem 1]. □

We define the Hamiltonian of the control problem (4.1) by

$$H(x, y, u, \varphi) := L(x, y, u) + \varphi f(x, y, u).$$

Note that the inequality constraint $\|u\|_0 \leq \tau$ is not taken into account in the Hamiltonian, which is different to the approach in Sect. 2. In addition, we defined the Hamiltonian in the qualified sense, that is, there is no "multiplier" $\varphi_0 \geq 0$ associated to the functional L by $\varphi_0 L$.

We are going to prove the maximum principle in integrated form first. The main difference to other works, e.g., [6, 14], is the construction of perturbations that satisfy the constraint $\|u\|_0 \leq \tau$. Here, we will adapt a result of [14] to generate these perturbations. It is based on Lyapunov's theorem.

Lemma 4.2 *Let $\rho \in (0, 1)$. Let $g_1, \ldots, g_m \in L^1(\Omega)$ be given. Then there is a sequence (E_ρ^n) of measurable subsets of Ω such that*

$$\int_{E_\rho^n} g_k \, dx = \rho \int_\Omega g_k \, dx \quad \forall k = 1, \ldots, m \; \forall n \in \mathbb{N}$$

and

$$\frac{1}{\rho} \chi_{E_\rho^n} \rightharpoonup^* 1 \quad \text{for } n \to \infty \text{ in } L^\infty(\Omega) = L^1(\Omega)^*.$$

Proof The proof is an adaptation of the proof of [14, Lemma 4.2]. It is included for the convenience of the reader.

Let (φ_n) be a dense subset of $L^1(\Omega)$. For $n \geq 0$, define $f^n : \Omega \to \mathbb{R}^{m+n}$ by

$$f^n = (g_1, \ldots, g_m, \varphi_1, \ldots, \varphi_n).$$

By the Lyapunov convexity theorem [7, Corollary IX.5], there is $E_\rho^n \subseteq \Omega$ such that $\rho \int_\Omega f^n \, dx = \int_{E_\rho^n} f^n \, dx$. By definition of f^n, this implies $\int_{E_\rho^n} g_k \, dx = \rho \int_\Omega g_k \, dx$ for all k.

Let now $\varphi \in L^1(\Omega)$ be given. Take $\epsilon > 0$. By density, there is N such that $\|\varphi - \varphi_N\|_{L^1(\Omega)} < \epsilon$. Then for all $n > N$, we get

$$\left| \int_\Omega \left(1 - \frac{1}{\rho} \chi_{E_\rho^n}\right) \varphi \, dx \right| \leq \left| \int_\Omega \left(1 - \frac{1}{\rho} \chi_{E_\rho^n}\right)(\varphi - \varphi_N) \, dx \right| + \left| \int_\Omega \left(1 - \frac{1}{\rho} \chi_{E_\rho^n}\right) \varphi_N \, dx \right|$$

$$\leq \frac{1 - \rho}{\rho} \epsilon + 0,$$

which proves the claim. $\qquad\square$

Corollary 4.3 *Let (E_ρ^n) be a sequence of measurable subsets of Ω such that*

$$\frac{1}{\rho} \chi_{E_\rho^n} \rightharpoonup^* 1 \quad \text{for } n \to \infty \text{ in } L^\infty(\Omega) = L^1(\Omega)^*.$$

Let $h \in L^2(\Omega)$ be given. Then $(1 - \frac{1}{\rho} \chi_{E_\rho^n}) h \to 0$ in $W^{-1,p}(\Omega)$ where $p \in (1, +\infty)$ for $d = 2$ and $p \in (1, 6)$ for $d = 3$.

Proof Due to the assumptions, we have $(1 - \frac{1}{\rho}\chi_{E_\rho^n})h \rightharpoonup 0$ in $L^2(\Omega)$. Under the conditions on p, the embedding $W_0^{1,p'}(\Omega) \hookrightarrow L^2(\Omega)$ is compact, where p' is given by $\frac{1}{p} + \frac{1}{p'} = 1$. Hence, the embedding $L^2(\Omega) \hookrightarrow W^{-1,p}(\Omega)$ is compact as well. $\quad\square$

Now we have all tools available to prove the maximum principle. The proof is very similar to the proofs in [6]. Hence, we will be brief on arguments that are similar to those in [6]. We first prove the maximum principle in integrated form.

Theorem 4.4 *Let $\bar{u}$ be a local solution of* (4.1a) *and* (4.1b) *in the $L^2(\Omega)$-sense with associated state $\bar{y} := y_{\bar{u}} \in W_0^{1,p'}(\Omega)$, where $p > d$ is such that $W_0^{1,p'}(\Omega) \hookrightarrow L^2(\Omega)$, where p' is given by $\frac{1}{p} + \frac{1}{p'} = 1$. Then there is $\bar{\varphi} \in W_0^{1,p'}(\Omega)$ that solves the adjoint equation*

$$A^*\bar{\varphi} = f_y(\cdot, \bar{y}, \bar{u})\bar{\varphi} + L_y(\cdot, \bar{y}, \bar{u}).$$

In addition,

$$\int_\Omega H(x, \bar{y}(x), \bar{u}(x), \bar{\varphi}(x))\,dx \le \int_\Omega H(x, \bar{y}(x), v(x), \bar{\varphi}(x))\,dx$$

for all $v \in L^2(\Omega)$ be with $\|v\|_0 \le \tau$.

Proof Let $v \in L^2(\Omega)$ with $\|v\|_0 \le \tau$. Set $h := f(\cdot, \bar{y}, v) - f(\cdot, \bar{y}, \bar{u})$, $m := 4$, and

$$(g_1, \ldots, g_m) := ((v - \bar{u})^2, |\bar{u}|_0, |v|_0, L(\cdot, \bar{y}, v) - L(\cdot, \bar{y}, \bar{u})).$$

Then by Lemma 4.2 and Corollary 4.3, for each $\rho > 0$ there is a set E_ρ such that

$$\int_{E_\rho} g_k\,dx = \rho \int_\Omega g_k\,dx \quad \forall k = 1, \ldots, m$$

and

$$\left\| \left(1 - \frac{1}{\rho}\chi_{E_\rho}\right)h \right\|_{W^{-1,p}(\Omega)} < \rho.$$

Let us set

$$u_\rho = \bar{u} + \chi_{E_\rho}(v - \bar{u}).$$

Then

$$\begin{aligned}
\|u_\rho\|_0 &= \|(1 - \chi_{E_\rho})\bar{u} + \chi_{E_\rho}v\|_0 \\
&= \|(1 - \chi_{E_\rho})\bar{u}\|_0 + \|\chi_{E_\rho}v\|_0 \\
&= \|\bar{u}\|_0 - \|\chi_{E_\rho}\bar{u}\|_0 + \|\chi_{E_\rho}v\|_0 = (1 - \rho)\|\bar{u}\|_0 + \rho\|v\|_0 \le \tau
\end{aligned}$$

and

$$\|u_\rho - \bar{u}\|_{L^2(\Omega)}^2 = \int_{E_\rho} (v - \bar{u})^2 \, \mathrm{d}x = \rho\|v - \bar{u}\|_{L^2(\Omega)}^2.$$

Hence, $J(\bar{u}) \le J(u_\rho)$ by local optimality of $\bar{u}$ for $\rho > 0$ small enough. Arguing as in [6, Lemma 2], we find

$$0 \le \lim_{\rho \searrow 0} \frac{1}{\rho}(J(u_\rho) - J(\bar{u})) = z^0,$$

where

$$z^0 = \int_\Omega L_y(x, \bar{y}(x), \bar{u}(x))z(x) + L(x, \bar{y}, v(x)) - L(x, \bar{y}(x), \bar{u}(x)) \, \mathrm{d}x$$

and $z \in W_0^{1,p}(\Omega)$ satisfies

$$Az = f_y(\cdot, \bar{y}, \bar{u})z + f(\cdot, \bar{y}, v) - f(\cdot, \bar{y}, \bar{u}).$$

In addition, there is $\bar{\varphi} \in W_0^{1,p'}(\Omega)$ [6, Theorem 2] that solves the adjoint equation

$$A^*\bar{\varphi} = f_y(\cdot, \bar{y}, \bar{u})\bar{\varphi} + L_y(\cdot, \bar{y}, \bar{u}).$$

This implies

$$
\begin{aligned}
0 \le z^0 &= \int_\Omega L(x, \bar{y}, v(x)) - L(x, \bar{y}(x), \bar{u}(x)) \, \mathrm{d}x \\
&\quad + \int_\Omega \bar{\varphi}(x)(f(x, \bar{y}(x), v(x)) - f(x, \bar{y}(x), \bar{u}(x))) \, \mathrm{d}x \\
&= \int_\Omega H(x, \bar{y}(x), v(x), \bar{\varphi}(x)) - H(x, \bar{y}(x), \bar{u}(x), \bar{\varphi}(x)) \, \mathrm{d}x,
\end{aligned}
$$

which is the claim. $\qquad\square$

Using Theorem 3.10 and the results of Sect. 3, we can turn the maximum principle from integrated to pointwise form.

Theorem 4.5 *Let $\bar{u}$ be a local solution of (4.1a)–(4.1b) in the $L^2(\Omega)$-sense with associated state $\bar{y} := y_{\bar{u}} \in W_0^{1,p}(\Omega)$, where $p > d$ is such that $W_0^{1,p'}(\Omega) \hookrightarrow L^2(\Omega)$ where p' is given by $\frac{1}{p} + \frac{1}{p'} = 1$, and adjoint $\bar{\varphi} \in W_0^{1,p'}(\Omega)$ given by Theorem 4.4.*

Then there is a number $s \le 0$ such that

$$s(\|\bar{u}\|_0 - \tau) = 0$$

and for almost all $x \in \Omega$

$$\bar{u}(x) = \underset{u \in \mathbb{R}}{\arg\min} \, H(x, \bar{y}(x), u, \bar{\varphi}(x)) + (-s)|u|_0.$$

In addition, we have the following properties for almost all $x \in \Omega$:

$$\bar{u}(x) \neq 0 \quad \Rightarrow \quad \bar{u}(x) = \underset{u\in\mathbb{R}}{\arg\min}\, H(x, \bar{y}(x), u, \bar{\varphi}(x)),$$

$$|\bar{u}(x)|_0 \cdot (\underset{u\in\mathbb{R}}{\inf}\, H(x, \bar{y}(x), u, \bar{\varphi}(x)) - H(x, \bar{y}(x), 0, \bar{\varphi}(x)) - s) \leq 0.$$

Proof Let us define g by

$$g(x, u) := H(x, \bar{y}(x), u, \bar{\varphi}(x)).$$

Then g is a normal integrand, and $g(\cdot, 0)$ is integrable. Due to Theorem 4.4, $\bar{u}$ solves

$$\min_{u\in\mathcal{U}_\tau\cap L^2(\Omega)} \int_\Omega g(x, u(x))\,\mathrm{d}x.$$

By Theorem 3.10, $\bar{u}$ solves (3.1). Hence, the results of Sect. 3 are applicable. Let $s \leq 0$ be as in Theorem 3.4. Then the claim follows with Corollaries 3.5, 3.7 and 3.8. $\qquad\qquad\square$

This result shows that the conditions (2.1) and (2.2), which we derived for an ODE control problem, are satisfied in adapted form in the PDE control problem.

5 Proximal gradient algorithm

In this section, we will analyze a proximal gradient algorithm applied to a problem with L^0 constraints. Here, we consider problems of the type

$$\min_{u\in\mathcal{U}_\tau\cap L^2(\Omega)} f(u) + \frac{\alpha}{2}\|u\|_{L^2(\Omega)}^2. \tag{5.1}$$

We are going to use the following set of assumptions.

Assumption 3

1. $\Omega \subseteq \mathbb{R}^d$ is Lebesgue measurable with $\mathrm{meas}(\Omega) \in (0, \infty)$, $\tau \in (0, \mathrm{meas}(\Omega))$.
2. The function $f : L^2(\Omega) \to \mathbb{R}$ is bounded from below and Fréchet differentiable. In addition, $\nabla f : L^2(\Omega) \to L^2(\Omega)$ is Lipschitz continuous with constant L_f, i.e.,

$$\|\nabla f(u_1) - \nabla f(u_2)\|_{L^2(\Omega)} \leq L_f \|u_1 - u_2\|_{L^2(\Omega)}$$

   holds for all $u_1, u_2 \in L^2(\Omega)$.
3. $\alpha \geq 0$.

These requirements on f are well-established in the context of first-order optimization methods. The requirement of global Lipschitz continuity of ∇f and knowledge of the Lipschitz modulus L_f can be overcome by a suitable back-tracking method, see [15, Section 3.3], which can be used in our situation as well.

Remark 5.1 Under some restrictions, the problem of Sect. 4 satisfies these assumptions. Let us assume that L is of the form $L(x, y, u) = L(x, y) + \frac{\alpha}{2}u^2$. Define $f(u) := \int_\Omega L(x, y_u(x)) \, dx$, where y_u is the solution of (4.1c). If the nonlinearity in the equation is linear in u, e.g., $f(x, y, u) = f(x, y) + u$, then f satisfies Assumption 3. See also the discussion in [13, Section 2.2].

Let us first prove a necessary optimality condition for (5.1). The proof is similar to Theorem 4.4 above.

Theorem 5.2 *Suppose f is a Fréchet differentiable mapping from $L^1(\Omega) \to \mathbb{R}$. Let $\bar{u}$ be a local solution of (5.1). Then it holds*

$$\int_\Omega \nabla f(\bar{u})\bar{u} \, dx + \frac{\alpha}{2}\|\bar{u}\|^2_{L^2(\Omega)} \leq \int_\Omega f(\bar{u})v \, dx + \frac{\alpha}{2}\|v\|^2_{L^2(\Omega)}$$

for all $v \in L^2(\Omega)$ be with $\|v\|_0 \leq \tau$.

In addition, there is a number $s \leq 0$ such that

$$s(\|\bar{u}\|_0 - \tau) = 0. \tag{5.2}$$

If $\alpha > 0$ then for almost all $x \in \Omega$ the following conditions are fulfilled:

$$\bar{u}(x) \neq 0 \quad \Rightarrow \quad \bar{u}(x) = -\frac{1}{\alpha}\nabla f(\bar{u})(x), \tag{5.3}$$

$$|\bar{u}(x)|_0 \cdot \left(-\frac{1}{2\alpha}|\nabla f(\bar{u})(x)|^2 - s\right) \leq 0. \tag{5.4}$$

If $\alpha = 0$ then $\nabla f(\bar{u}) = 0$.

Proof Let us set $F(u) := f(u) + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)}$, which is Fréchet differentiable on $L^2(\Omega)$ with gradient $\nabla F(u) = \nabla f(u) + \alpha u$. Let $v \in L^2(\Omega)$ with $\|v\|_0 \leq \tau$. Set $m := 5$, and

$$(g_1, \ldots, g_m) := ((v - \bar{u})^2, |\bar{u}|_0, |v|_0, \nabla F(\bar{u}), |v - \bar{u}|).$$

Then by Lemma 4.2, for each $\rho > 0$ there is a set E_ρ such that $\int_{E_\rho} g_j \, dx = \rho \int_\Omega g_j \, dx$ for all $j = 1 \ldots m$. As in the proof of Theorem 4.4, the function $u_\rho := \bar{u} + \chi_{E_\rho}(v - \bar{u})$ satisfies $\|u_\rho\|_0 \leq \tau$ and $\|u_\rho - \bar{u}\|^2_{L^2(\Omega)} = \rho\|v - \bar{u}\|^2_{L^2(\Omega)}$. Due to Fréchet differentiability and the construction of E_ρ and u_ρ, we have

$$F(u_\rho) - F(\bar{u}) = \nabla F(\bar{u})(u_\rho - \bar{u}) + o(\|u_\rho - \bar{u}\|_{L^1(\Omega)}) + \frac{\alpha}{2}\|u_\rho - \bar{u}\|^2_{L^2(\Omega)}$$

$$= \rho\nabla F(\bar{u})(v - \bar{u}) + o(\rho) + \rho\frac{\alpha}{2}\|v - \bar{u}\|^2_{L^2(\Omega)}.$$

Dividing by $\rho > 0$ and passing to the limit $\rho \searrow 0$, implies by local optimality

$$0 \le \nabla F(\bar{u})(v - \bar{u}) + \frac{\alpha}{2}\|v - \bar{u}\|^2_{L^2(\Omega)}$$
$$= \nabla f(\bar{u})(v - \bar{u}) + \frac{\alpha}{2}\|v\|^2_{L^2(\Omega)} - \frac{\alpha}{2}\|\bar{u}\|^2_{L^2(\Omega)},$$

which proves the first claim. The second claim follows from Theorem 3.4 and Corollaries 3.5 and 3.7 with $\tilde{v} = -\frac{1}{2\alpha}|\nabla f(\bar{u})|^2$.   □

Let us briefly give the motivation of the proximal gradient algorithm. The well-known steepest descent method applied to the unconstrained differentiable problem $\min_u f(u)$ amounts to the iteration

$$u_{k+1} = u_k - t_k \nabla f(u_k), \tag{5.5}$$

where $t_k > 0$ is a suitable step-size. It is immediate that u_{k+1} is a solution of the unconstrained problem

$$\min_u f(u_k) + \nabla f(u_k) \cdot (u - u_k) + \frac{1}{2t_k}\|u - u_k\|^2_{L^2(\Omega)}. \tag{5.6}$$

While it is impossible to add the constraint $\|u\|_0 \le \tau$ to the iteration procedure (5.5), this constraint can be easily imposed on the problem (5.6). The resulting proximal gradient (or forward-backward) algorithm reads as follows. Here, we replaced the parameter t_k by a fixed parameter L, which takes the place of $\frac{1}{t_k}$.

Algorithm 1 (Proximal gradient algorithm) Choose $L > 0$ and $u_0 \in L^2(\Omega)$. Set $k = 0$.

1. Compute u_{k+1} as solution of

$$\min_{u \in \mathcal{U}_\tau \cap L^2(\Omega)} f(u_k) + \nabla f(u_k) \cdot (u - u_k) + \frac{L}{2}\|u - u_k\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)} \tag{5.7}$$

2. Set $k := k + 1$, go to step 1.

The functional to be minimized in (5.7) can be written as an integral functional $\int_\Omega g(x, u(x)) \, dx$ with g defined by

$$g(x, u) = f(u_k) + \nabla f(u_k)(x) \cdot (u - u_k(x)) + \frac{L}{2}(u - u_k(x))^2 + \frac{\alpha}{2}u^2.$$

The pointwise minimum of g is realized by the function $\tilde{u} \in L^2(\Omega)$ defined by

$$\tilde{u}(x) := \frac{Lu_k(x) - \nabla f(u_k)(x)}{L + \alpha}.$$

Clearly, $g(\cdot, 0)$ is integrable, Assumption 1 is satisfied, and the results of Sect. 3 are applicable. Hence, a solution of (5.7) can be computed as in Theorem 3.4. Here, $L > 0$ is important: note that integrability of $g(\cdot, u)$ implies $u \in L^2(\Omega)$. It is easy to verify that

$$g(x, \tilde{u}(x)) - g(x, 0) = -\frac{1}{2(L + \alpha)} \left(Lu_k(x) - \nabla f(u_k)(x) \right)^2,$$

which corresponds to $\tilde{v}$ in Theorem 3.4.

Lemma 5.3 *Let Assumption 3 be satisfied. Let $L > 0$ and $u_k \in L^2(\Omega)$ be given. Then (5.7) is solvable. In addition, there is $\lambda_{k+1} \geq 0$ such that for every solution u_{k+1} of (5.7) it holds*

$$\lambda_{k+1}(\|u_{k+1}\|_0 - \tau) = 0,$$

and u_{k+1} solves

$$\min_{u \in L^2(\Omega)} f(u_k) + \nabla f(u_k) \cdot (u - u_k) + \frac{L}{2} \|u - u_k\|_{L^2(\Omega)}^2 + \frac{\alpha}{2} \|u\|_{L^2(\Omega)}^2 + \lambda_{k+1} \|u\|_0.$$

Moreover, for almost all $x \in \Omega$ we have

$$u_{k+1}(x) \neq 0 \quad \Rightarrow \quad |u_{k+1}(x)| \geq \sqrt{\frac{2\lambda_{k+1}}{L + \alpha}} \tag{5.8}$$

and

$$|u_{k+1}(x)|_0 \cdot \left(-\frac{1}{2(L + \alpha)} \left(Lu_k(x) - \nabla f(u_k)(x) \right)^2 + \lambda_{k+1} \right) \leq 0. \tag{5.9}$$

Proof Existence of solutions follows from Theorem 3.4. The properties of $\lambda_{k+1} := -s$, where s is as in Theorem 3.4, are consequences of Corollary 3.7 and Corollary 3.8. The claim (5.8) is a consequence of Corollary 3.8 and [15, Corollary 3.9]. Finally, Corollary 3.5 implies (5.9). ∎

The iterates of the algorithm satisfy the following properties.

Theorem 5.4 *Let Assumption 3 be satisfied. Suppose $L > L_f$. Let (u_k) be a sequence of iterates generated by Algorithm 1. Then it holds that*:

1. *The sequences (u_k) and $(\nabla f(u_k))$ are bounded in $L^2(\Omega)$ if $\alpha > 0$.*
2. *The sequence $(f(u_k) + \frac{\alpha}{2} \|u_k\|_{L^2(\Omega)})$ is monotonically decreasing and converging.*
3. $\sum_{k=0}^{\infty} \|u_{k+1} - u_k\|_{L^2(\Omega)}^2 < \infty.$

Proof These claims can be proven as in [15, Theorem 3.13]. ∎

Let us define the following sequence

$$\chi_k(x) := |u_k(x)|_0.$$

Using Eq. (5.8), we have the following estimate of (χ_k), which is similar to [15, Lemma 3.12].

Lemma 5.5 *Let* (u_k) *be iterates of Algorithm* 1. *Then it holds*

$$\|u_{k+1} - u_k\|_{L^2(\Omega)}^2 \geq \frac{2\min(\lambda_k, \lambda_{k+1})}{L + \alpha}\|\chi_{k+1} - \chi_k\|_{L^1(\Omega)}.$$

Proof Let $x \in \Omega$ such that $\chi_{k+1}(x) \neq \chi_k(x)$. Then $|\chi_{k+1}(x) - \chi_k(x)| = 1$, and exactly one of $u_{k+1}(x)$ and $u_k(x)$ is zero. Suppose $u_{k+1}(x) = 0$ and $u_k(x) \neq 0$. Then $|u_{k+1}(x) - u_k(x)| = |u_k(x)| \geq \sqrt{\frac{2\lambda_k}{L+\alpha}}$. If $u_{k+1}(x) \neq 0$ and $u_k(x) = 0$ then $|u_{k+1}(x)| \geq \sqrt{\frac{2\lambda_{k+1}}{L+\alpha}}$. And the claim is proven. $\square$

Under the assumption that (λ_k) is bounded from below by a positive number, we can prove feasibility of weak limit points of the algorithm. In the general situation, it is not clear how to prove such a result as the map $u \mapsto \|u\|_0$ is not weakly sequentially lower semi-continuous from $L^2(\Omega) \to \mathbb{R}$.

Theorem 5.6 *Let Assumption* 3 *be satisfied. Suppose* $L > L_f$. *Let* (u_k) *be a sequence of iterates generated by Algorithm* 1. *Suppose*

$$\liminf_{k\to\infty} \lambda_k > 0,$$

where (λ_k) *is as in Lemma* 5.3.

Then $\chi_k \to \bar{\chi}$ *in* $L^1(\Omega)$, *and every weak sequential limit point* $\bar{u}$ *of* (u_k) *is feasible for the* L^0 *constraint, i.e.* $\|\bar{u}\|_0 \leq \tau$.

Proof Let $\lambda := \liminf_{k\to\infty} \lambda_k > 0$. Then for all k sufficiently large, we have

$$\|u_{k+1} - u_k\|_{L^2(\Omega)}^2 \geq \frac{\lambda}{L + \alpha}\|\chi_{k+1} - \chi_k\|_{L^1(\Omega)}.$$

The summability of $\|u_{k+1} - u_k\|_{L^2(\Omega)}^2$ implies those of $\|\chi_{k+1} - \chi_k\|_{L^1(\Omega)}$. Hence (χ_k) is a Cauchy sequence in $L^1(\Omega)$, $\chi_k \to \bar{\chi}$ in $L^1(\Omega)$, and $\bar{\chi}$ is a characteristic function. As (χ_k) is trivially bounded in $L^\infty(\Omega)$, it follows $\chi_k \to \bar{\chi}$ in $L^p(\Omega)$ for all $p < \infty$.

Let now (u_{k_n}) be a subsequence with $u_{k_n} \rightharpoonup \bar{u}$ in $L^2(\Omega)$. Let $\varphi \in L^\infty(\Omega)$. Since $\chi_k(x) := |u_k(x)|_0$, we have $\int_\Omega (1 - \chi_{k_n})u_{k_n}\varphi \, \mathrm{d}x = 0$ for all n. Passing to the limit in this equation yields $\int_\Omega (1 - \bar{\chi})\bar{u}\varphi \, \mathrm{d}x = 0$. Since $\varphi \in L^\infty(\Omega)$ was arbitrary, this implies $(1 - \bar{\chi})\bar{u} = 0$ almost everywhere, which in turn implies $|\bar{u}|_0 \leq \bar{\chi}$ almost everywhere, as both functions $\bar{\chi}$ and $|u|_0$ only attain the values 0 and 1. And it follows

$$\|\bar{u}\|_0 \leq \|\bar{\chi}\|_{L^1(\Omega)} = \lim_{k\to\infty} \|\chi_k\|_{L^1(\Omega)} = \lim_{k\to\infty} \|u_k\|_0 \leq \tau,$$

and $\bar{u}$ is feasible for the L^0 constraint. $\square$

Moreover, we can prove strong convergence under additional assumptions on ∇f. See also the related result [15, Theorem 3.18]. Here, we assume that ∇f maps weakly to strongly converging sequences.

Theorem 5.7 *Let Assumption* 3 *be satisfied. Suppose* $L > L_f$. *Let us assume complete continuity of* ∇f *from* $L^2(\Omega)$ *to* $L^2(\Omega)$, *i.e., for all sequences* (v_k) *in* $L^2(\Omega)$ *the following implication*

$$v_k \rightharpoonup v \text{ in } L^2(\Omega) \;\Rightarrow\; \nabla f(v_k) \to \nabla f(v) \text{ in } L^2(\Omega) \tag{5.10}$$

holds. In addition, we require $\alpha > 0$.

Let (u_k) *be a sequence of iterates generated by Algorithm* 1. *Suppose*

$$\liminf_{k \to \infty} \lambda_k > 0, \tag{5.11}$$

where (λ_k) *is as in Lemma* 5.3.

Then $u_{k_n} \rightharpoonup \bar{u}$ *in* $L^2(\Omega)$ *implies* $u_{k_n} \to \bar{u}$ *in* $L^2(\Omega)$. *In addition, for almost all* $x \in \Omega$ *the following condition is fulfilled*

$$\bar{u}(x) \neq 0 \quad\Rightarrow\quad \bar{u}(x) = -\frac{1}{\alpha}\nabla f(\bar{u})(x).$$

Proof If $u_{k+1}(x) \neq 0$ then $\alpha u_{k+1}(x) = -(\nabla f(u_k)(x) + L(u_{k+1}(x) - u_k(x)))$. This implies

$$\chi_{k+1} u_{k+1} = -\chi_{k+1}\frac{1}{\alpha}(\nabla f(u_k) + L(u_{k+1} - u_k)).$$

Adding the equation $(1 - \chi_{k+1})u_{k+1} = 0$, yields

$$\alpha u_{k+1} = -\chi_{k+1}(\nabla f(u_k) + L(u_{k+1} - u_k)). \tag{5.12}$$

Let now $u_{k_n} \rightharpoonup \bar{u}$ in $L^2(\Omega)$. Then $\nabla f(u_{k_n}) \to \nabla f(\bar{u})$ in $L^2(\Omega)$ by complete continuity of ∇f. In addition, $u_{k+1} - u_k \to 0$ in $L^2(\Omega)$ by Theorem 5.4. The right-hand side in (5.12) converges strongly in $L^2(\Omega)$ by Lemma 5.8 below, which implies the strong convergence $u_{k_n} \to \bar{u}$ in $L^2(\Omega)$. In addition, in the limit we obtain from (5.12) $\alpha \bar{u} = -\bar{\chi}\nabla f(\bar{u})$. $\qquad\square$

Let us compare the properties of limit points $\bar{u}$ with the necessary optimality conditions (5.2)–(5.4) according to Theorem 5.2. The above result only proves the implication (5.3) for limit points $\bar{u}$. It seems to be impossible to prove the remaining two conditions (5.2) and (5.4). An obvious choice for s in those formulas would be any limit point of $(-\lambda_k)$. Under the assumption (5.11), we would get $s < 0$. However, it seems impossible to prove $\|\bar{u}\|_0 = \tau$: the mapping $u \mapsto |u|_0$ is merely lower semicontinuous at $u = 0$, so that we can only prove $\|\bar{u}\|_0 \leq \tau$. And it is not clear that the complementarity condition (5.2) is satisfied in the limit. In order to prove (5.4), a natural idea would be to pass to the limit in the condition (5.9). At best we can expect to get

$$|\bar{u}(x)|_0 \cdot \left(-\frac{1}{2(L+\alpha)}(L\bar{u}(x) - \nabla f(\bar{u})(x))^2 - s\right) \leq 0.$$

This is different to (5.4) because of the presence of the prox-parameter $L > 0$ in the inequality.

We close the section with the following auxiliary result, whis was used in the proof of Theorem 5.7. Note that an application of Hölder inequality implies strong convergence in $L^p(\Omega)$ only for $p < 2$.

Lemma 5.8 *Let* meas$(\Omega) < \infty$. *Let sequences* (χ_k) *and* (g_k) *be given such that* $\|\chi_k\|_{L^\infty(\Omega)} \leq 1$, $\chi_k \to \bar{\chi}$ *in* $L^1(\Omega)$, *and* $g_k \to \bar{g}$ *in* $L^2(\Omega)$. *Then* $\chi_k g_k \to \bar{\chi}\bar{g}$ *in* $L^2(\Omega)$.

Proof The sequences admit pointwise a.e. converging subsequences (g_{k_n}), (χ_{k_n}) together with a dominating function $a \in L^2(\Omega)$ with $|g_{k_n}| \leq a$, see [3, Theorem 4.9]. Then $\chi_{k_n} g_{k_n} \to \bar{\chi}\bar{g}$ in $L^2(\Omega)$ by dominated convergence. A subsequence-subsequence argument finishes the proof. □

Data availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Aubin, J.-P., Frankowska, H.: Set-Valued Analysis. Vol. 2. Systems & Control: Foundations & Applications. Birkhäuser Boston, Inc., Boston (1990). https://doi.org/10.1007/978-0-8176-4848-0
2. Berliocchi, H., Lasry, J.-M.: Intégrandes normales et mesures paramétrées en calcul des variations. Bulletin de la Société Mathématique de France **101**, 129–184 (1973). https://doi.org/10.24033/bsmf.1755
3. Brezis, H.: Functional Analysis, Sobolev Spaces and Partial Differential Equations. Universitext. Springer, New York, pp. xiv+599 (2011)
4. Buttazzo, G., Maiale, F.P., Velichkov, B.: Shape optimization problems in control form. Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl. **32**(3), 413–435 (2021). https://doi.org/10.4171/rlm/942

5. Casas, E., Raymond, J.-P., Zidani, H.: Optimal control problem governed by semilinear elliptic equations with integral control constraints and pointwise state constraints. In: Control and estimation of distributed parameter systems (Vorau, 1996). Vol. 126. Internat. Ser. Numer. Math. Birkhäuser, Basel, pp. 89–102 (1998)

6. Casas, E.: Pontryagin's principle for optimal control problems governed by semilinear elliptic equations. In: Control and Estimation of Distributed Parameter Systems: Nonlinear Phenomena (Vorau, 1993). Vol. 118. Internat. Ser. Numer. Math. Birkhäuser, Basel, pp. 97–114 (1994)

7. Diestel, J., Uhl Jr., J.J.: Vector measures. Mathematical Surveys, No. 15. With a foreword by B. J. Pettis. American Mathematical Society, Providence, pp. xiii+322 (1977)

8. Gotoh, J.-Y., Takeda, A., Tono, K.: DC formulations and algorithms for sparse optimization problems. Math. Program. Ser. B **169**(1), 141–176 (2018). https://doi.org/10.1007/s10107-017-1181-0

9. Ioffe, A.D., Tihomirov, V.M.: Theorie der Extremalaufgaben. Translated from the Russian by Bernd Luderer. VEB Deutscher Verlag der Wissenschaften, Berlin, p. 399 (1979)

10. Ito, K., Kunisch, K.: Optimal control with $L^p(\Omega)$, $p \in [0, 1)$, control cost. SIAM J. Control Optim. **52**(2), 1251–1275 (2014). https://doi.org/10.1137/120896529

11. Kalise, D., Kunisch, K., Sturm, K.: Optimal actuator design based on shape calculus. Math. Models Methods Appl. Sci. **28**(13), 2667–2717 (2018). https://doi.org/10.1142/S0218202518500586

12. Kanzow, C., Raharja, A.B., Schwartz, A.: An augmented Lagrangian method for cardinality-constrained optimization problems. J. Optim. Theory Appl. **189**(3), 793–813 (2021). https://doi.org/10.1007/s10957-021-01854-7

13. Natemeyer, C., Wachsmuth, D.: A proximal gradient method for control problems with non-smooth and non-convex control cost. Comput. Optim. Appl. **80**(2), 639–677 (2021). https://doi.org/10.1007/s10589-021-00308-0

14. Raymond, J.P., Zidani, H.: Pontryagin's principle for state-constrained control problems governed by parabolic equations with unbounded controls. SIAM J. Control Optim. **36**(6), 1853–1879 (1998). https://doi.org/10.1137/S0363012996302470

15. Wachsmuth, D.: Iterative hard-thresholding applied to optimal control problems with $L^0(\Omega)$ control cost. SIAM J. Control Optim. **57**(2), 854–879 (2019). https://doi.org/10.1137/18M1194602