A DIFFERENCE OF CONVEX FUNCTIONS APPROACH FOR
SPARSE PDE OPTIMAL CONTROL PROBLEMS WITH NONCONVEX
COSTS

PEDRO MERINO‡

ABSTRACT. We propose a local regularization of elliptic optimal control problems which
involves the nonconvex $L^q$ fractional penalizations in the cost function. The proposed
Huber type regularization allows us to formulate the PDE constrained optimization for-
mulation as a DC programming problem (difference of convex functions) that is useful to
obtain necessary optimality conditions and tackle its numerical solution by applying the
well known DC algorithm used in nonconvex optimization problems. By this procedure
we approximate the original problem in terms of a consistent family of parameterized
nonsmooth problems for which there are efficient numerical methods available. Finally,
we present numerical experiments to illustrate our theory with different configurations
associated to the parameters of the problem.

1. INTRODUCTION

Several optimal control problems governed by PDEs with sparse solutions have been
considered in recent years. One of the pioneer works on this subject c.f.
[23] introduced
optimal control problems with $L^1$–norm penalization in order to promote sparse optimal
solutions. These solutions are characterized by having small supports, which are inter-
preted as a “localized” action of the optimal control. This particular feature of sparse
optimal controls is relevant in applications because it is rather difficult in practice imple-
menting optimal controls taking values on the whole domain, which is the usual case of
optimal control problems with the usual $L^2$–term in its cost functional.

Another interesting class of optimal control problems involving sparsity were considered
in [4] and [3] where the set of feasible controls is chosen in the space of regular Borel
measures. Therefore, optimal controls can be supported in a set of zero Lebesgue measure.
A complete review on this subject, including parabolic problems, can be found in [5] and
the references therein.

A less explored approach that offers sparse solutions induced by a penalization term was
considered in [19] which refers to penalizations consisting in nonconvex $L^q$–functionals with
$q \in [0, 1)$. These kind of penalizations have many important applications, for instance: in
inverse problems on the reconstruction of the sparsest solution in undetermined systems
[21], image restoration [13], compressive sensing [12] and optimal control problems [19].

In particular, the $L^0$–functional is a difficult problem which corresponds to the selection
of the most representative variables of the optimization process, extending the notion of
cardinality of the control variable in finite dimensions, represented by the $\ell^0$ norm, which
is well known to be an NP–hard problem. $L^q$–functionals with $q \in (0, 1)$ on the other
hand, are a natural approximation to $L^0$–functionals. However, they are neither convex
nor differentiable.

2010 Mathematics Subject Classification. 90C26, 90C46, 49J20, 49K20.

This research has been supported by Research Project PIJ-15-26 funded by Escuela Politécnica Na-
cional, Quito–Ecuador. Moreover, we acknowledge partial support of SENESCYT–MATHAmSud project
SOCDE “Sparse Optimal Control of Differential Equations”.
In [19] a similar problem is considered involving a penalization term for the control variable involving the $H^1_0$–norm. This allows to get and explicit optimality system that can be solved directly by semi–smooth Newton methods. In our case, we consider a Tikhonov term in the $L^2$–norm. Although existence of optimal controls can be argued in this case, uniqueness of the solution is not expected as shown in a simple example below.

Due to the lack of convexity and differentiability these type of costs are difficult to tackle numerically. In this paper, we address the numerical solution of this type of problems by regularizing the fractional $L^q$–functionals; for this purpose, we introduce a Huber–like smoothing function which regularizes the nonconvex $L^q$ term. In this way, we obtain a family of regularized nonsmooth problems whose objective functional can be expressed as a DC-function (“DC” stands for difference of convex functions), which reveals the underlying convexity of this class of problems. Although the regularized problem remains nonconvex and nondifferentiable, we can take advantage of the DC structure of the functional by applying known tools from the convex analysis and DC programming theories in order to derive optimality conditions and prove that the regularization is consistent. Moreover, we propose a numerical method based on the DC-Algorithm (DCA). It follows that the proposed DC splitting leads to a primal–dual updating that only requires the numerical resolution of a convex $L^1$–norm penalized optimal control problem in each iteration, for which there are efficient numerical methods.

It is worth to mention that although our methodology is proposed for elliptic problems, it can be extended for different boundary conditions, parabolic problems or optimal control problems involving other type of equations.

This paper is organized as follows. In section 1 we introduce the non convex optimal control problems endowed with $L^q$–functionals with $q = \frac{1}{p}$, and $p > 1$. In Section 2 we introduce a Huber–like smoothing function in order to regularize the nonconvex optimal control problems. We show that the regularized problems can be expressed as a difference of convex functions and derive optimality conditions in Section 3. The box–constrained case is discussed at the end of this section. In addition, we provide a proof that the solution of the regularized version of the optimal control problem approximates its solution when the regularized parameter tends to infinity. Section 4 is devoted to the numerical solution by proposing a DC–Algorithm based method. We finish this research by showing numerical examples and numerical evidence of the efficiency of the proposed method.

1.1. Setting of the problem. For $p > 1$, let us define the mapping $\Upsilon_p : L^2(\Omega) \to \mathbb{R}$ defined by

$$u \mapsto \Upsilon_p(u) := \int_{\Omega} |u|^{\frac{1}{p}}.$$

Let $\Omega$ a bounded Lipschitz domain in $\mathbb{R}^n$ ($n = 2$ or $n = 3$) with boundary $\Gamma$. We are interested in the following optimal control problem with a $L^{1/p}$–fractional penalization term [1]. Therefore, for $\alpha > 0$ and $\beta > 0$ we consider the optimal control problem:

$$\min_{(y,u) \in L^2(\Omega) \times H^1_0(\Omega)} \left\{ \frac{1}{2} \|y - y_d\|^2_{L^2(\Omega)} + \alpha \frac{1}{2} \|u\|^2_{L^2(\Omega)} + \beta \Upsilon_p(u) \right\}$$

subject to:

$$Ay = u + f, \quad \text{in } \Omega,$$

$$y = 0, \quad \text{on } \Gamma,$$
where $f$ is a given function in $L^2(\Omega)$ and $A$ is a uniformly elliptic second order differential operator of the form

$$(2) \quad (Ay)(x) = -\sum_{i,j=1}^{n} \frac{\partial}{\partial x_i} \left( a_{ij}(x) \frac{\partial y(x)}{\partial x_j} \right) + c_0 y(x).$$

Here, the coefficients $a_{ij} \in C^{0,1}(\Omega)$, and $c_0 \in L^\infty(\Omega)$. Moreover, the matrix $(a_{ij})_{i,j}$ is symmetric and fulfill the uniform ellipticity condition:

$$\exists \sigma > 0 : \sum_{i,j=1}^{n} a_{ij}(x) \xi_i \xi_j \geq \sigma |\xi|^2, \quad \forall \xi \in \mathbb{R}^n, \text{ for almost all } x \in \Omega.$$

We will denote the adjoint of $A$ by $A^\ast$. Moreover, associated to the elliptic operator $A$, we define the bilinear form

$$a(y, v) := \sum_{i,j=1}^{n} a_{ij}(x) \frac{\partial y(x)}{\partial x_i} \frac{\partial v(x)}{\partial x_j} + c_0 y(x)v(x),$$

which we use to define the problem:

$$(3) \quad a(y, v) = (w, v), \quad \forall v \in H^1_0(\Omega).$$

It is well known that (3) has a unique solution belonging to the space $H^1_0(\Omega)$. Let $S : L^2(\Omega) \to H^1_0(\Omega)$ be the linear and continuous operator which assigns to every $v \in L^2(\Omega)$ the corresponding solution $y = y(w) \in H^1_0(\Omega)$ satisfying (3). Thus, the state equation: $Ay = u$ in $\Omega$, with homogeneous Dirichlet boundary conditions considered in (P), is understood in the weak sense of equation (3). In this way, the state $y$ associated to the control $u$ has the representation $y = S(u + f)$, which in turn, allows us to formulate the usual reduced problem:

$$(P') \quad \min_{u \in L^2(\Omega)} J(u) := \frac{1}{2} \|Su + Sf - y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2} \|u\|^2_{L^2(\Omega)} + \beta \Upsilon_p(u).$$

**Theorem 1.** There exists a solution $\bar{u} \in L^2(\Omega)$ for the reduced problem $(P')$.

We postpone the proof of this result to Section 4 where we prove that a sequence of solutions of approximating problems of the form $\min_u J_\gamma(u)$, converges to the solution of $(P')$.

**Remark 1.** The question of uniqueness is more delicate. The following example of the minimization of a real function has two solutions. Let $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = \frac{1}{2}(x - a)^2 + \beta|x|^\frac{1}{2}$. By choosing $a = 1 + \frac{1}{2}$ and $\beta = 1$, it is easy to verify that $f$ has two minimum points at $x_1 = 0$ and $x_2 = 1$ with the minimum value $f(0) = f(1) = \frac{9}{8}$. Therefore, we cannot expect uniqueness of the solution for problem $(P')$ in view of nonconvexity of cost function.

Following the work of Stadler [23] where $L^1$–norm penalization optimal control problems are considered, we expect that some analogous properties also hold for problem $(P)$. For example, it is expected that a local solution for $(P)$ vanishes if the parameter $\beta$ is large enough. We address to this question in the following lemma.

**Lemma 1.** Let $S^\ast$ be the adjoint operator of $S$, and let $M > 0$. If $\beta \geq \beta_0$ with $\beta_0 = M^\frac{1}{n+1} \|S^\ast(Sf - y_d)\|_{L^\infty(\Omega)}$, then problem $(P)$ has a local minimum at $\bar{u} = 0$ in $B_\infty(0, M)$ with associated state $\bar{y} = Sf$.  


Proof. Taking into account the reduced form \( (P') \), we argue analogously to \cite[Lemma 3.1]{2020}. Let us take \( u \in B_{\infty}(0, M) \), then \( |u(x)| < M \) for almost all \( x \) in \( \Omega \). Computing the difference of the cost values we have:

\[
J(u) - J(0) = \frac{1}{2} \| Su + Sf - y_d \|_{L^2(\Omega)}^2 + \alpha \frac{1}{2} \| u \|_{L^2(\Omega)}^2 + \beta \Upsilon_p(u)
\]

\[
= \frac{1}{2} \| Su \|_{L^2(\Omega)}^2 + (Su, Sf - y_d)_{L^2(\Omega)} + \alpha \frac{1}{2} \| u \|_{L^2(\Omega)}^2 + \beta \Upsilon_p(u),
\]

\[
\geq \frac{1}{2} \| Su \|_{L^2(\Omega)}^2 - \| u \|_{L^1(\Omega)} \| S^*(Sf - y_d) \|_{L^\infty(\Omega)} + \alpha \frac{1}{2} \| u \|_{L^2(\Omega)}^2 + \beta \Upsilon_p(u),
\]

\[
\geq \int_{\Omega} \beta |u|^\frac{1}{p} - \| u \| \| S^*(Sf - y_d) \|_{L^\infty(\Omega)} \, dx,
\]

\[
\geq \int_{\Omega} \beta_0 |u|^\frac{1}{p} - \| u \| \| S^*(Sf - y_d) \|_{L^\infty(\Omega)} \, dx.
\]

By the definition of \( \beta_0 \) it follows that

\[
J(u) - J(0) \geq \int_{\Omega} \left( M^{\frac{p-1}{p}} - |u|^{\frac{p-1}{p}} \right) |u|^\frac{1}{p} \| S^*(Sf - y_d) \|_{L^\infty(\Omega)} \, dx > 0,
\]

where the nonnegativity is obtained by our assumption \( u \in B_{\infty}(0, M) \).


2. HUBER–TYPE REGULARIZATION OF THE OPTIMAL CONTROL PROBLEM

In order to analyze problem \((P)\) we formulate a family of regularized problems, by means of the following Huber–type regularization of the absolute value. Extending the classical Huber \( C^1 \) regularization of the absolute value, we propose a Huber regularization \( \Upsilon_{p,\gamma} \) which takes into account the fractional powers defining \( \Upsilon_p \). The resulting function to the power \( 1/p \) is a locally convex regularization for the nonconvex and non differentiable term, see Figure 1 below. For \( \gamma \gg 1 \), we define

\[
h_{p,\gamma}(v) = \begin{cases} 
\frac{\gamma^{p-1}}{p} |v|^p, & \text{if } v \in [-\frac{1}{\gamma}, \frac{1}{\gamma}], \\
|v| + \frac{1-\frac{1}{p}}{\gamma}, & \text{otherwise.}
\end{cases}
\]

Remark 2. The function \( h_{p,\gamma} \) is a local regularization of the absolute value for different smoothing polynomial powers. In addition, notice that by construction, we have the relation

\[
h_{p,\gamma}(v) \leq |v|, \quad \forall v \in \mathbb{R}.
\]

It is worth to notice that \((4)\) is different from the local regularization proposed in \cite[pg. 1971 eq.(5.1)]{1971} which majorizes \( \Upsilon_p(u) \). Both regularization terms can be used to compute upper and lower bounds for the cost functions of \((P)\), respectively. Although they may appear similar, observe that \((4)\) approximates \( g \) nonsmoothly in a neighborhood of 0. This fact is crucial to express our objective functional as a difference of convex functions. In fact, the representation as a DC–function is not possible using the regularization proposed by \cite{1971}. Therefore, by using the Huber–type regularization we are able to approximaze \((P)\) by sequence of \( L^1 \)–sparse problems. The resulting DC–algorithm will be introduced in Section 4.
Now, we have the basic tool in order to formulate a regularized version of \((P)\). We introduce the function \(\Upsilon_{p,\gamma}\) defined by
\[
  u \mapsto \Upsilon_{p,\gamma}(u) := \int_\Omega h_{p,\gamma}(u(x))^{\frac{1}{p}} \, dx.
\]
The regularized problem is obtained by replacing \(\Upsilon_p\) by \(\Upsilon_{p,\gamma}\). Therefore, the surrogate problem reads:
\[
(P_{\gamma}) \quad \begin{cases}
  \min_{(y,u)} \frac{1}{2} \|y - y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2} \|u\|^2_{L^2(\Omega)} + \beta \Upsilon_{p,\gamma}(u) \\
  \text{subject to:} \\
  Ay = u + f \quad \text{in } \Omega, \\
  y = 0 \quad \text{on } \Gamma.
\end{cases}
\]

Now, we proceed to formulate the reduced optimal control problem from \((P_{\gamma})\) by replacing the control–to–state operator \(S\). Let \(f\) be the regular part of the functional, which is \(f(u) = \frac{1}{2} \|Su - y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2} \|u\|^2_{L^2(\Omega)}\). Now we have the reduced problem,
\[
(7) \quad \min_u J_{\gamma}(u) := f(u) + \beta \Upsilon_{p,\gamma}.
\]
From [19, Lemma 5.1] it is known that if a sequence \((u_n)\) such that \(u_n \to u\) in \(L^1(\Omega)\) then \(\Upsilon_p(u_n) \to \Upsilon_p(u)\) as \(n \to \infty\). In the case of \(\Upsilon_{\gamma,p}\) we have the following continuity property.

**Lemma 2.** Let \((u_n)\) be a sequence such that \(u_n \to u\) in \(L^1(\Omega)\), then

\[ \Upsilon_{p,\gamma}(u_n) \to \Upsilon_{p,\gamma}(u), \quad \text{when } n \to \infty, \]

for all \(p > 1\) and all \(\gamma > 0\).

Proof. Analogously to [19, Lemma 5] we define the following sets:

\[
\begin{align*}
\Omega_{n,1} &= \{ x : |u(x)| \leq \frac{1}{\gamma} \text{ and } |u_n(x)| \leq \frac{1}{\gamma} \}, \\
\Omega_{n,2} &= \{ x : |u(x)| > \frac{1}{\gamma} \text{ and } |u_n(x)| > \frac{1}{\gamma} \}, \\
\Omega_{n,3} &= \{ x : |u(x)| \leq \frac{1}{\gamma} \text{ and } |u_n(x)| > \frac{1}{\gamma} \} \cup \{ x : |u(x)| > \frac{1}{\gamma} \text{ and } |u_n(x)| \leq \frac{1}{\gamma} \},
\end{align*}
\]

which we use to estimate \(\left| \int_{\Omega} h_{p,\gamma}(u(x))^\frac{1}{p} - h_{p,\gamma}(u_n(x))^\frac{1}{p} \, dx \right|\) according to (4). By our assumption \(u_n \to u\) in \(L^1(\Omega)\) whenever \(n \to \infty\), in \(\Omega_{n,1}\) we have that

\[
\left| \int_{\Omega_{n,1}} h_{p,\gamma}(u(x))^\frac{1}{p} - h_{p,\gamma}(u_n(x))^\frac{1}{p} \, dx \right| \leq \left( \frac{\gamma^{p-1}}{p} \right)^\frac{1}{p} \int_{\Omega_{n,1}} |u(x)| - |u_n(x)| \, dx,
\]

(8)

\[
\leq \left( \frac{\gamma^{p-1}}{p} \right)^\frac{1}{p} \int_{\Omega} |u(x) - u_n(x)| \, dx \to 0.
\]

Now, in \(\Omega_{n,2}\) we can estimate

\[
\left| \int_{\Omega_{n,2}} h_{p,\gamma}(u(x))^\frac{1}{p} - h_{p,\gamma}(u_n(x))^\frac{1}{p} \, dx \right| \leq \int_{\Omega_{n,2}} \left( |u(x)| + \frac{1}{\gamma} \right)^\frac{1}{p} - \left( |u_n(x)| + \frac{1}{\gamma} \right)^\frac{1}{p} \, dx
\]

\[
\leq \int_{\Omega_{n,2}} |u(x)| - |u_n(x)| \, dx,
\]

\[
\leq \int_{\Omega_{n,2}} |u(x) - u_n(x)| \, dx.
\]

By applying Hölder inequality in the last integral, and by our convergence assumption we have

\[
\left| \int_{\Omega_{n,2}} h_{p,\gamma}(u(x))^\frac{1}{p} - h_{p,\gamma}(u_n(x))^\frac{1}{p} \, dx \right| \leq |\Omega|^{\frac{p}{p+\gamma}} \left( \int_{\Omega} |u(x) - u_n(x)| \, dx \right)^\frac{1}{p}
\]

\[
\to 0.
\]

(9)

Finally, we estimate in \(\Omega_{n,3}\). Without loss of generality we assume that \(\{ x : |u(x)| \leq \frac{1}{\gamma} \text{ and } |u_n(x)| > \frac{1}{\gamma} \}\). The neglected part can be argued in the same way by interchanging the role of \(|u(x)|\) and \(|u_n(x)|\). Taking into account that the relation: \(|u(x)| \leq 1/\gamma < |u_n(x)|\) is fulfilled in \(\Omega_{n,3}\), it follows that

\[
\left( \frac{\gamma^{p-1}}{p} \right) |u(x)|^p < |u_n(x)| + \frac{1}{\gamma} \frac{1-p}{p},
\]
follows:

\[
\left| \int_{\Omega_{n,3}} h_{p,\gamma}(u(x)) \frac{1}{p} - h_{p,\gamma}(u_n(x)) \frac{1}{p} \, dx \right| \leq \int_{\Omega_{n,3}} \left| h_{p,\gamma}(u(x)) - h_{p,\gamma}(u_n(x)) \right| \frac{1}{p} \, dx \\
= \int_{\Omega_{n,3}} \left| \left( \frac{\gamma^{p-1}}{p} \right) |u(x)|^p - |u_n(x)| - \frac{1}{\gamma} \frac{1-p}{p} \right| \frac{1}{p} \, dx \\
= \int_{\Omega_{n,3}} \left( |u_n(x)| + \frac{1}{\gamma} \frac{1-p}{p} - \left( \frac{\gamma^{p-1}}{p} \right) |u(x)|^p \right) \frac{1}{p} \, dx.
\]

(10)

Furthermore, in \( \Omega_{n,3} \) we have that \( \frac{1}{\gamma} < |u_n(x)| + \frac{1-p}{\gamma} < |u_n(x)| \), from which we obtain that

\[
|u_n(x)| + \frac{1}{\gamma} \frac{1-p}{p} < |u_n(x)|^p \left( \frac{\gamma^{p-1}}{p} \right).
\]

By replacing (11) in (10) we get the following relation

\[
\left| \int_{\Omega_{n,3}} h_{p,\gamma}(u(x)) \frac{1}{p} - h_{p,\gamma}(u_n(x)) \frac{1}{p} \, dx \right| \leq \left( \frac{\gamma^{p-1}}{p} \right) \frac{1}{p} \int_{\Omega_{n,3}} \left( |u_n(x)|^p - |u(x)|^p \right) \frac{1}{p} \, dx,
\]

\[
= \left( \frac{\gamma^{p-1}}{p} \right) \frac{1}{p} \int_{\Omega_{n,3}} \left| |u_n(x)|^p - |u(x)|^p \right| \frac{1}{p} \, dx,
\]

(12)

By applying the mean value theorem, we have that there is a \( \xi(x) \) such that \( |u(x)| < \xi(x) < |u_n(x)| \) for almost all \( x \) in \( \Omega_{n,3} \) that satisfies \( |u_n(x)|^p - |u(x)|^p = p|\xi(x)|^{p-1}(|u_n(x)| - |u_n(x)|) \). Hence, using this relation and applying Hölder inequality we have

\[
\int_{\Omega_{n,3}} \left| |u_n(x)|^p - |u(x)|^p \right| \frac{1}{p} \, dx \leq \int_{\Omega_{n,3}} \left( p \frac{1}{p} |\xi(x)|^{\frac{p-1}{p}} \right) \left| |u_n(x)| - |u(x)| \right| \frac{1}{p} \, dx \leq p \frac{1}{p} \int_{\Omega_{n,3}} |\xi(x)| \, dx \int_{\Omega_{n,3}} \left| |u_n(x)| - |u(x)| \right| \, dx.
\]

Thereby, the right–hand side of (12) tends to 0 as \( n \to 0 \). Finally, collecting estimates (8), (9) and (12) the result of the lemma is proved.

**Lemma 3.** \( J_\gamma \) converges to \( J \) uniformly as \( \gamma \to \infty \).

Proof. We argue the uniform convergence of \( J_\gamma \) to \( J \) by using the definition of the Huber regularization (4). Since \( J_\gamma \) and \( J \) differ on the nonconvex term, we analyze the difference \( |\Upsilon_{p,\gamma}(u) - \Upsilon_p(u)| \) in the sets \( \Omega_\gamma = \{ x \in \Omega : |u(x)| \leq \frac{1}{\gamma} \} \) and \( \Omega_\gamma^c = \{ x \in \Omega : |u(x)| > \frac{1}{\gamma} \} \) as follows:

\[
\left| \int_{\Omega} h_{p,\gamma}(u(z))^{\frac{1}{p}} - |u(z)|^{\frac{1}{p}} \, dz \right| \leq \int_{\Omega_\gamma} \left| h_{p,\gamma}(u(z))^{\frac{1}{p}} - |u(z)|^{\frac{1}{p}} \right| \, dz + \int_{\Omega_\gamma} \left| h_{p,\gamma}(u(z))^{\frac{1}{p}} - |u(z)|^{\frac{1}{p}} \right| \, dz \\
\leq \int_{\Omega_\gamma} \left( \frac{\gamma^{\frac{1}{p-1}}}{p} |u|^{\frac{p-1}{p}} - |u|^{\frac{1}{p}} \right) \, dz + \int_{\Omega_\gamma} \left( \frac{1}{\gamma} \frac{1-p}{p} \right)^{\frac{1}{p}} - |u|^{\frac{1}{p}} \, dz.
\]
Using the fact that $|u| \leq \frac{1}{\gamma}$ in $\Omega$, we have
\[
\left| \int_{\Omega} h_{p, \gamma}(u)^{\frac{1}{p}} - |u|^{\frac{1}{p}} \, dx \right| \leq \int_{\Omega} \frac{1}{\gamma \frac{p}{p^*}} + \frac{1}{\gamma} \, dx + \frac{1}{\gamma \frac{p}{p^*}} \int_{\Omega} \frac{1 - p}{p} |u|^{\frac{1}{p}} \, dx,
\]
where the last terms clearly tend to 0 as $\gamma \to \infty$. Moreover, this limit does not depend on $u$, implying the uniform convergence of $\Upsilon_{p, \gamma}$ to $\Upsilon_p$.

The next theorem addresses the question about existence of a solution of problem $(P_\gamma)$.

**Theorem 2.** There exists a solution $\bar{u}_\gamma \in L^2(\Omega)$ for the regularized problem $(P_\gamma)$.

Proof. Existence of a solution can be argued by standard techniques from the direct methods of calculus of variations, see for instance [7]. Owing to the relation:
\[
J_\gamma(u) \geq \frac{\alpha}{2} \|u\|_{L^2(\Omega)}^2 \quad \text{for all} \quad u \in L^2(\Omega),
\]
we have that $J_\gamma$ is coercive in $L^2(\Omega)$. Moreover, from Lemma 2 it follows that $J_\gamma(u) = f(u) + \Upsilon_{p, \gamma}(u)$ is continuous in $L^1(\Omega)$ for every $p > 1$ and every $\gamma \gg 1$, and because the embedding $L^2(\Omega) \hookrightarrow L^1(\Omega)$ it is also continuous in $L^2(\Omega)$. Therefore, by Theorem 1.15 in [7] it follows that $J_\gamma$ has a minimum in $L^2(\Omega)$ which, in what follows, will be denoted by $\bar{u}_\gamma$.

### 3. Optimality Conditions of the Regularized Problem

Our aim in this section is deriving an optimality system for problem $(P_\gamma)$ via a DC–programming approach. As mentioned earlier, the key idea is introducing an $L^1$–norm penalization which allows us to formulate our problem as a minimization of a difference of convex functions, with functions $G$ and $H$ such that:

\[
J_\gamma(u) = G(u) - H(u).
\]

A function that can be expressed in this form is known as a DC–function and several problems involving this type of functions have been analyzed and its theory can be found in the monograph of Hiriart Urruty [15] or in [10].

Let us focus on how to express the cost function of problem $(P_\gamma)$ as a convenient difference of convex functions and then rely on the theory of DC programming. We start by introducing the following quantity, which will be frequently used throughout this paper:

\[
\delta = \frac{\gamma^{\frac{p-1}{p}}}{p^*}.
\]

The next step is to define $G$ and $H$ in (13) as follows:

\[
G : \quad L^2(\Omega) \to \mathbb{R}, \quad u \mapsto G(u) := \frac{1}{2} \|Su + Sf - y_d\|_{L^2(\Omega)}^2 + \alpha \|u\|_{L^2(\Omega)}^2 + \beta \delta \|u\|_{L^1(\Omega)},
\]

\[
H : \quad L^2(\Omega) \to \mathbb{R}, \quad u \mapsto H(u) := \beta \left( \delta \|u\|_{L^1(\Omega)} - \Upsilon_{p, \gamma}(u) \right).
\]
Lemma 4. The real function $j : \mathbb{R} \to \mathbb{R} \cup \{0\}$, defined by

$$
(16) \quad j(z) = \begin{cases} 
\delta |z| - (|z| + \frac{1}{\gamma} \frac{1-p}{p})^{\frac{1}{p}}, & \text{if } |z| > \frac{1}{\gamma} \\
0, & \text{if } |z| \leq \frac{1}{\gamma},
\end{cases}
$$

is nonnegative, convex and continuously differentiable and its derivative is given by

$$
(17) \quad j'(z) = \begin{cases} 
\delta \text{sign}(z) - \frac{1}{p} \left(|z| + \frac{1}{\gamma} \frac{1-p}{p}\right)^{\frac{1-p}{p}} \text{sign}(z), & \text{if } |z| > \frac{1}{\gamma} \\
0, & \text{if } |z| \leq \frac{1}{\gamma}.
\end{cases}
$$

Proof. Let us first check differentiability. It is clear that $j$ is differentiable if $|z| < \frac{1}{\gamma}$ or $|z| > \frac{1}{\gamma}$, where $j'(z) = 0$ and $j'(z) = \delta \text{sign}(z) - \frac{1}{p} \left(|z| + \frac{1}{\gamma} \frac{1-p}{p}\right)^{\frac{1-p}{p}} \text{sign}(z)$, respectively. Therefore, we check differentiability at $z = \pm \frac{1}{\gamma}$. Consider $z = -\frac{1}{\gamma}$, since $j(\pm \frac{1}{\gamma}) = 0$ and $|\frac{1}{\gamma} + h| < \frac{1}{\gamma}$ for sufficiently small $h$, we have that

$$
\lim_{h \to 0^+} \frac{j(z + h) - j(z)}{h} = \lim_{h \to 0^+} \frac{j(-\frac{1}{\gamma} + h)}{h}
$$

for sufficiently small $h$. On the other hand, since $-\frac{1}{\gamma} + h < 0$ for sufficiently small $h$

$$
\lim_{h \to 0^-} \frac{j(z + h) - j(z)}{h} = \lim_{h \to 0^-} \frac{j(-\frac{1}{\gamma} + h)}{h}
$$

where we apply the binomial theorem to get

$$
\lim_{h \to 0^-} \frac{\left(\frac{1}{\gamma p}\right)^{\frac{1}{p}} - \delta h - \left(\frac{1}{\gamma p} - h\right)^{\frac{1}{p}}}{h} = \lim_{h \to 0^-} \frac{\left(\frac{1}{\gamma p}\right)^{\frac{1}{p}} - \delta h - \left(\frac{1}{\gamma p} + \frac{1-p}{p}\right)^{\frac{1}{p}} h + o(h)}{h} = 0.
$$

Therefore $j'(-\frac{1}{\gamma}) = 0$. Analogously, it also follows that $j'(\frac{1}{\gamma}) = 0$, which implies formula (17). Moreover, an straightforward observation reveals that $j'$ is continuous, therefore $j$ is continuously differentiable. Convexity follows by noticing that function $\mathbb{R}_+ \ni z \mapsto (z + \frac{1}{\gamma} \frac{1-p}{p})^{1/p}$ is concave because it is the composition of an affine function and a concave function. Thus, for $z > \frac{1}{\gamma}$, we find that the function

$$
\mathbb{R}_+ \ni z \mapsto \delta z - \left(z + \frac{1}{\gamma} \frac{1-p}{p}\right)^{1/p},
$$

is convex and monotone increasing which, by composition with the absolute value, implies the convexity of $j$. Finally, we make the simple but important observation that $j$ vanishes in the interval $[-\frac{1}{\gamma}, \frac{1}{\gamma}]$. This, together with the convexity of $j$, implies that $j$ is nonnegative.

Now, by employing the function $j$ we can write $H$ as follows:

$$
H : \mathbb{L}^2(\Omega) \to \mathbb{R}
$$

$$
(u \mapsto H(u) = \int_{\Omega} j(u) dx).
$$
Lemma 5. The functions $G$ and $H$ defined in (15) are convex.

Proof. Since $\alpha \geq 0$ and $\beta \geq 0$, it is clear that function $G$ is strictly convex if $\beta + \alpha > 0$. In the case of $H$, convexity follows from Lemma 4.

Having defined the functions $H$ and $G$, it is clear that the representation (13) of $J_\gamma$ has been set up. Therefore, $J_\gamma$ is a DC-function and we can express optimality conditions in terms of $G$ and $H$ by considering the following formulation for problem (7):

\[(DC) \quad \min_u J_\gamma(u) = G(u) - H(u),\]

Lemma 6. The function $H$ defined in (15) is Gâteaux differentiable, and its derivative $\delta H(\bar{u}; \cdot)$ is represented by $(\beta \bar{w}, \cdot)$, where $\bar{w} \in L^2(\Omega)$ depends on $\bar{u}$, $p$ and $\gamma$, and it is given by

\[
(19) \quad \bar{w}(x) := \begin{cases} \\
\delta - \frac{1}{p} \left( |\bar{u}(x)| + \frac{1}{\gamma} \frac{1 - p}{p} \right)^{\frac{1-p}{p}} \text{sign}(\bar{u}(x)), & \text{if } |\bar{u}(x)| > \frac{1}{\gamma}, \\
0, & \text{otherwise.}
\end{cases}
\]

Proof. First, notice that $j'(z)$ given in (17) satisfies that

\[
(20) \quad 0 < |j'(z)| = \left| \delta - \frac{1}{p} \left( |z| + \frac{1}{\gamma} \frac{1 - p}{p} \right)^{\frac{1-p}{p}} \right| < \delta, \quad \text{for } |z| > \frac{1}{\gamma}.
\]

Therefore, by using (20) and the properties of $j$ established in Lemma 4, we apply [11, Theorem 2.7, pg. 19] in order to deduce that superposition operator $u \mapsto j(u)$ is Gâteaux differentiable from $L^2(\Omega)$ into $L^2(\Omega)$, and its Gâteaux derivative in the direction $v$ is given by $j'(u)v \in L^2(\Omega)$. Hence, Theorem 7.4-1 in [6] allows us to compute the Gâteaux derivative of $H$ at $\bar{u}$ in any direction $v \in L^2(\Omega)$ by

\[
(21) \quad \delta H(\bar{u}, v) = \int_{\Omega} j'(\bar{u}(x)) vdx = (\bar{w}, v),
\]

with $\bar{w}$ given by (19).

3.1. First–order necessary conditions. The following part of this paper moves on describing the derivation of first order necessary optimality conditions for problem (P$_\gamma$). The conditions for local and global optimality can be found in [15, Proposition 3.1 and 3.2] or in [11]. We will use the following well known result from DC–programming theory, which permits the characterization of local minima.

Proposition 1. Let $G$ and $H$, the convex functions defined in (15). If $\bar{u}$ is a local minimum of the DC–function $J_\gamma = G - H$, then $\bar{u}$ satisfies the following critical point condition:

\[
(22) \quad \partial H(\bar{u}) \subset \partial G(\bar{u}).
\]

The next result establishes an optimality system with the help of the last proposition.

Theorem 3. Let $\bar{u}$ a solution of (P$_\gamma$), then there exist $\bar{y} = S\bar{u}$ in $H^1_0(\Omega)$, an adjoint state $\phi \in H^1_0(\Omega)$, a multiplier $\zeta \in L^2(\Omega)$ and $\bar{w}$ given by (19) such that the following optimality
system is satisfied:

\[(23a)\]
\[A \bar{y} = \bar{y} + f, \quad \text{in } \Omega,\]
\[\bar{y} = 0, \quad \text{on } \Gamma,\]

\[(23b)\]
\[A^* \bar{\phi} = \bar{y} - y_d, \quad \text{in } \Omega,\]
\[\bar{\phi} = 0, \quad \text{on } \Gamma,\]

\[(23c)\]
\[\bar{\phi} + \alpha \bar{u} + \beta (\delta \zeta - \bar{w}) = 0,\]
\[\zeta(x) = 1, \quad \text{if } \bar{u}(x) > 0,\]
\[\zeta(x) = -1, \quad \text{if } \bar{u}(x) < 0,\]
\[|\zeta(x)| \leq 1, \quad \text{if } \bar{u}(x) = 0,\]
\[\text{for almost all } x \in \Omega.\]

Proof. Clearly, equation (23a) is equivalent to \(Su = \bar{y}\). By standard properties of the subdifferential calculus c.f.\cite{16}, the subdifferential of \(G\) at \(\bar{u}\) is given by \(\partial G(\bar{u}) = \nabla f(\bar{u}) + \beta \delta \partial \| \cdot \|_{L^1(\Omega)}(\bar{u})\). By Lemmas \(2\) and \(6\) it follows that \(\partial H(\bar{u})\) consists in the singleton \(\{\bar{w}\}\). Thus, condition (22) becomes

\[(24)\]
\[\bar{w} \in \nabla f(\bar{u}) + \beta \delta \partial \| \cdot \|_{L^1(\Omega)}(\bar{u}).\]

Since \(S\) is a linear and continuous operator from \(L^2(\Omega)\) into \(L^2(\Omega)\), the computation of \(\nabla f(\bar{u})\) is straightforward, see for instance \(8\). Therefore, for \(u \in L^2(\Omega)\) we have that

\[(25)\]
\[\nabla f(\bar{u})u = (Su, S\bar{u} + Sf - y_d)_{L^2(\Omega)} + \alpha(u, \bar{u})_{L^2(\Omega)}\]
\[= (u, \alpha \bar{u} + S^*(\bar{y} - y_d))_{L^2(\Omega)}.\]

Moreover, by introducing the adjoint state \(\bar{\phi} \in H_0^1(\Omega)\) as the solution of the adjoint equation: \((23b)\)

\[A^* \bar{\phi} = \bar{y} - y_d, \quad \text{in } \Omega,\]
\[\bar{\phi} = 0, \quad \text{on } \Gamma,\]

we are able to write \(\bar{\phi} = S^*(\bar{y} - y_d)\) \((S^*\) denoting the adjoint control-to–state operator).

On the other hand, it is well known \cite{17} Chapter 0.3.2, that any \(\zeta \in \partial \| \cdot \|_{L^1(\Omega)}(\bar{u})\) is characterized by

\[(26)\]
\[\zeta(x) \begin{cases} 
1, & \text{if } \bar{u}(x) > 0, \\
-1, & \text{if } \bar{u}(x) < 0, \\
\in [-1,1], & \text{if } \bar{u}(x) = 0.
\end{cases}\]

In this way, from (25) we obtain that \(\nabla f(\bar{u}) = \bar{\phi} + \alpha \bar{u}\) which together with (26) imply the existence of \(\zeta \in \partial \| \cdot \|_{L^1(\Omega)}(\bar{u}) \subset L^\infty(\Omega)\) which allows us to write (24) in the form:

\[(27)\]
\[\bar{\phi} + \alpha \bar{u} + \beta (\delta \zeta - \bar{w}) = 0.\]

An important question regarding the regularized problem \((P_\gamma)\) is about the convergence of the solutions of \((P_\gamma)\) to a solution of the original problem \((P)\) when \(\gamma \to \infty\). We address this question in the following Theorem.

**Theorem 4.** Let \(\{\bar{u}_\gamma\}_{\gamma}\) a sequence of solutions of problem \((P_\gamma)\). There exists a subsequence in \(L^2(\Omega)\) whose weak limit, denoted by \(u^*\), is a solution for problem \((P)\).
Proof. We begin by noticing that the sequence \((\bar{u}_\gamma)_{\gamma > 0}\) is bounded in \(L^2(\Omega)\). Indeed, since \(S_0 = 0\), optimality of \(\bar{u}_\gamma\) results in
\[
\frac{\alpha}{2} \|u_\gamma\|_{L^2(\Omega)}^2 \leq J_\gamma(\bar{u}_\gamma) \leq J_\gamma(0) = \frac{1}{2} \|y_d\|_{L^2(\Omega)}^2,
\]
which implies the boundedness of \((\bar{u}_\gamma)_{\gamma > 0}\) in \(L^2(\Omega)\) for \(\alpha > 0\).

As usual, reflexivity of \(L^2(\Omega)\) allows us to extract a weakly convergent subsequence, denoted again by \((\bar{u}_\gamma)_{\gamma > 0}\) which has the limit \(u^* \in L^2(\Omega)\). Arguing that the optimality of \(\bar{u}_\gamma\) implies that
\[
J_\gamma(\bar{u}_\gamma) \leq J_\gamma(u) \quad \text{for any } u \in L^2(\Omega) \text{ and taking into account (5), it follows that}
\]
\[
J_\gamma(\bar{u}_\gamma) \leq J_\gamma(u) = \frac{1}{2} \|y - y_d\|_{L^2(\Omega)}^2 + \frac{\alpha}{2} \|u\|_{L^2(\Omega)}^2 + \beta \Upsilon_p(u) = J(u),
\]
implying
\[
(28) \quad \liminf_{\gamma \to 0} J_\gamma(\bar{u}_\gamma) \leq J(u).
\]
Furthermore, we also have that
\[
\liminf_{\gamma \to 0} J_\gamma(\bar{u}_\gamma) = \liminf_{\gamma \to \infty} \left( f(\bar{u}_\gamma) + \Upsilon_{p,\gamma}(\bar{u}_\gamma) \right)
\geq \liminf_{\gamma \to \infty} f(\bar{u}_\gamma) + \liminf_{\gamma \to \infty} \Upsilon_{p,\gamma}(\bar{u}_\gamma)
\geq f(u^*) + \liminf_{\gamma \to \infty} \Upsilon_{p,\gamma}(\bar{u}_\gamma),
\]
where the last relation is obtained in view of the weakly lower semicontinuity of \(f\). Now, we turn to the last term in (29). Let us consider the sets: \(\Omega_\gamma = \{x \in \Omega : |u_\gamma(x)| \leq \frac{1}{2}\gamma\}\) and \(\Omega^c_\gamma = \{x \in \Omega : |u_\gamma(x)| > \frac{1}{2}\gamma\}\). By definition of the Huber–type regularization \([4]\) we find that
\[
\liminf_{\gamma \to \infty} \Upsilon_{p,\gamma}(\bar{u}_\gamma) = \liminf_{\gamma \to \infty} \int_{\Omega} (h_{p,\gamma}(\bar{u}_\gamma))^{\frac{1}{p}} dx
\geq \liminf_{\gamma \to \infty} \int_{\Omega} \chi_{\Omega_\gamma} \frac{\gamma^{\frac{p-1}{p}}}{p^{\frac{1}{p}}} |\bar{u}_\gamma| dx + \liminf_{\gamma \to \infty} \int_{\Omega} \chi_{\Omega^c_\gamma} \left( |\bar{u}_\gamma| + \frac{1 - p}{\gamma} \right) \frac{1}{p}^{\frac{1}{p}} dx.
(30)
\]
In \(\Omega_\gamma\) we known that \(|\bar{u}_\gamma| \leq \frac{1}{2}\gamma\), therefore, \(\chi_{\Omega_\gamma} \frac{\gamma^{\frac{p-1}{p}}}{p^{\frac{1}{p}}} |\bar{u}_\gamma| \leq \gamma^{-\frac{1}{p}} p^{-\frac{1}{p}}\). Hence, the first term in (30) vanishes as \(\gamma \to \infty\). Now, we drive our analysis to the last term in (30). Consider the function \(g : \mathbb{R}_+ \to \mathbb{R}\) defined by \(x \mapsto x^{\frac{1}{p}}\). It is clear that \(g\) is a concave differentiable function for \(p > 1\). In addition, it is known that \(g\) satisfies the inequality
\[
g(x + h) \geq g(x) - g'(x)h.
(31)
\]
Applying this inequality we have
\[
\liminf_{\gamma \to \infty} \int_{\Omega} \Theta_{\gamma} \left( |\bar{u}_{\gamma}| + \frac{1}{\gamma} \frac{1}{p} \right)^{\frac{1}{p}} \, dx
\]
\[
\geq \liminf_{\gamma \to \infty} \int_{\Omega} \Theta_{\gamma} \left( |\bar{u}_{\gamma}|^\frac{1}{p} - \frac{1}{\gamma} \frac{1}{p} \left( |\bar{u}_{\gamma}(x)| + \frac{1}{\gamma} \frac{1}{p} \right)^{\frac{1}{p}} \right) \, dx
\]
\[
= \liminf_{\gamma \to \infty} \int_{\Omega} \bar{u}_{\gamma}^\frac{1}{p} \, dx - \limsup_{\gamma \to \infty} \int_{\Omega} \Theta_{\gamma} \left( |\bar{u}_{\gamma}(x)| + \frac{1}{\gamma} \frac{1}{p} \right)^{\frac{1}{p}} \, dx
\]
(32) \quad = \liminf_{\gamma \to \infty} \int_{\Omega} \bar{u}_{\gamma}^\frac{1}{p} \, dx,

where the last relation follows since the integrant of the right term is bounded hence the second term vanishes. Taking into account (32) and (30) and by the uniform convergence of \( \bar{\Upsilon}_{p,\gamma} \) see Lemma 3 we have that \( \bar{\Upsilon}_{p,\gamma} \) converges to \( \bar{\Upsilon}_p \), see [7]. Therefore, we arrive to the inequality:

\[
\liminf_{\gamma \to \infty} \Upsilon_{p,\gamma}(\bar{u}_{\gamma}) \geq \liminf_{\gamma \to \infty} \int_{\Omega} |\bar{u}_{\gamma}|^\frac{1}{p} \, dx = \int_{\Omega} |\bar{u}|^\frac{1}{p} \, dx,
\]
which, combined with (29) implies that

\[
\liminf_{\gamma \to \infty} J_\gamma(\bar{u}_{\gamma}) \geq f(u^*) + \int_{\Omega} |\bar{u}|^\frac{1}{p} \, dx = J(u^*).
\]

Finally, (34) and (28) imply that \( J(u^*) = \liminf_{\gamma \to \infty} J_\gamma(\bar{u}_{\gamma}) \).

**Remark 3.** Here, we have proved that the weak limit of a subsequence of solutions of \( \mathcal{P}_\gamma \) is a solution of problem \( \mathcal{P} \) in \( L^2(\Omega) \). Thus, Theorem 7 is proved. In case we have a penalization term in the \( H^1_0(\Omega) \)-norm as in [19], it follows by the compact embedding of \( H^1_0(\Omega) \) into \( L^2(\Omega) \) we can extract a subsequence that converges strongly in \( L^2(\Omega) \) and by the continuity of \( \Upsilon_{p,\gamma} \) we obtain the this result in this with strong convergence in \( L^2(\Omega) \).

### 3.2. First–order necessary conditions with box–constraints.

Since box–constraints are important in applications, we give a further discussion when they are included in the optimal control problem [7]. Let us consider the set of feasible controls given by:

\[
U_{ad} = \{ u \in L^2(\Omega) : u_a(x) \leq u(x) \leq u_b(x), \text{ a.a. } x \in \Omega \},
\]
where \( u_a \) and \( u_b \) are given functions in \( L^\infty(\Omega) \) satisfying \( u_a(x) < u_b(x) \) a.a. \( x \in \Omega \).
The control constrained optimal control problem reads:

\[
\begin{align*}
\min_{(y,u)} & \frac{1}{2} \|y - y_d\|_{L^2(\Omega)}^2 + \frac{\alpha}{2} \|u\|_{L^2(\Omega)}^2 + \beta \Upsilon_p(u) \\
\text{subject to:} & \quad u \in U_{ad} \quad \text{and} \quad Ay = u + f, \quad \text{in } \Omega, \\
& \quad y = 0, \quad \text{on } \Gamma.
\end{align*}
\]

\((P_C)\)

Remark 4. It follows by definition [35] that \(U_{ad} \subset B_\infty(0, M)\) with \(M = \max\{\|u_a\|_{L^\infty(\Omega)}, \|u_b\|_{L^\infty(\Omega)}\}\). Therefore, according to Lemma 1 if \(\beta > \beta_0 = M \frac{\alpha}{\beta} \|S^*(Sf - y_d)\|_{L^\infty(\Omega)}\) then \(\bar{u} = 0\) is solution of \((P_C)\).

Analogous to the unconstrained optimal control problem \((P')\), after introducing the control–to–state operator \(S\) and replacing \(\Upsilon_p\) by \(\Upsilon_{p,\gamma}\), we introduce the regularized control constrained problem

\[
\begin{align*}
\min_{(y,u)} & \frac{1}{2} \|y - y_d\|_{L^2(\Omega)}^2 + \frac{\alpha}{2} \|u\|_{L^2(\Omega)}^2 + \beta \Upsilon_{p,\gamma}(u) \\
\text{subject to:} & \quad u \in U_{ad} \quad \text{and} \quad Ay = u + f, \quad \text{in } \Omega, \\
& \quad y = 0, \quad \text{on } \Gamma.
\end{align*}
\]

\((P_{C,\gamma})\)

define a DC representation of the cost functional for problem \((P_C)\) by including the indicator function \(\chi_{U_{ad}}\) for the admissible control set:

\[
\begin{align*}
G : \ L^2(\Omega) & \rightarrow \mathbb{R} \\
 u & \mapsto G(u) := \frac{1}{2} \|Su + Sf - y_d\|_{L^2(\Omega)}^2 + \alpha \|u\|_{L^2(\Omega)}^2 + \beta \delta \|u\|_{L^1(\Omega)} + I_{U_{ad}}, \\
H : \ L^2(\Omega) & \rightarrow \mathbb{R} \\
 u & \mapsto H(u) := \beta \left(\delta \|u\|_{L^1(\Omega)} - \Upsilon_{p,\gamma}(u)\right).
\end{align*}
\]

Thus, by similar arguments as in the unconstrained case and taking into account that \(\partial I_{U_{ad}}(u)\) corresponds to the normal cone of \(U_{ad}\) at \(\bar{u}\), we can derive an analogous optimality system.

Theorem 5. Let \(\bar{u}\) a solution of \((P_{C,\gamma})\), then there exist \(\bar{y} = S\bar{u}\) in \(H^1_0(\Omega)\), an adjoint state \(\phi \in H^1_0(\Omega)\) and a multiplier \(\zeta \in L^2(\Omega)\) and \(\bar{\omega}\) given by (19) such that the following optimality system is satisfied:

\[
\begin{align*}
A\bar{y} & = \bar{u} + f \quad \text{in } \Omega, \\
\bar{y} & = 0 \quad \text{on } \Gamma, \\
A^*\bar{\phi} & = \bar{y} - y_d \quad \text{in } \Omega, \\
\bar{\phi} & = 0 \quad \text{on } \Gamma, \\
\langle \bar{\phi} + \alpha \bar{u} + \beta (\delta \zeta - \bar{\omega}), u - \bar{u} \rangle & \geq 0, \quad \forall u \in U_{ad} \\
\zeta(x) & = 1, \quad \text{si } \bar{u}(x) > 0, \\
\zeta(x) & = -1, \quad \text{si } \bar{u}(x) < 0, \\
|\zeta(x)| & \leq 1, \quad \text{si } \bar{u}(x) = 0,
\end{align*}
\]

for almost all \(x \in \Omega\).
Moreover, there exist $\lambda_a$ and $\lambda_b$ in $L^2(\Omega)$ such that the last optimality system can be written as a KKT optimality system:

$$
\begin{align*}
\bar{A} \bar{y} &= \bar{u} + f \quad \text{in } \Omega, \\
\bar{y} &= 0 \quad \text{on } \Gamma,
\end{align*}
$$

(38a)

$$
\begin{align*}
\bar{A^*} \bar{\phi} &= \bar{y} - y_d \quad \text{in } \Omega, \\
\bar{\phi} &= 0 \quad \text{on } \Gamma,
\end{align*}
$$

(38b)

$$
\bar{\phi} + \alpha \bar{u} + \beta (\delta \zeta - \bar{w}) + \lambda_b - \lambda_a = 0
$$

(38c)

$$
\begin{align*}
\lambda_a &\geq 0, \quad \lambda_b \geq 0, \\
\lambda_a (\bar{u} - u_a) &= 0, \quad \lambda_b (u_b - \bar{u}) = 0,
\end{align*}
$$

(38d)

$$
\begin{align*}
\zeta(x) &= 1 \quad \text{si } \bar{u}(x) > 0, \\
\zeta(x) &= -1 \quad \text{si } \bar{u}(x) < 0, \\
|\zeta(x)| &\leq 1 \quad \text{si } \bar{u}(x) = 0,
\end{align*}
$$

(38e)

Proof. This theorem is proved by following the arguments of the proof of Theorem 3, where variational inequality (37c) follows by taking into consideration classical results on convex analysis and the fact that $\bar{w} \in \nabla f(\bar{u}) + \beta \delta \partial \| \cdot \|_{L^1(\Omega)}(\bar{u}) + \partial I_{U_{ad}}(\bar{u})$.

In addition, by the usual projection operator $P_{U_{ad}}$ (see [14, Lemma 1.11]) on the admissible control set, the variational inequality (37c) can be equivalently rewritten in equation form:

$$
\bar{u} = P_{U_{ad}} \left[ -\frac{1}{\alpha} (\bar{\phi} + \beta (\delta \zeta - \bar{w})) \right].
$$

(39)

4. Numerical solution via the DC Algorithm (DCA)

In the former section we have derived necessary optimality conditions for problem ($P_\gamma$) and problem ($P_{C\gamma}$), which are suitable for applying the Semi-Smooth Newton method (SSN). However, SSN does not guarantee descent of the objective function.

By the nature of our problem we turn our attention to its numerical solution by adapting the DC algorithm. The application of DC algorithm to our problem leads to a numerical scheme which relies on numerical methods for solving sparse $L^1$ optimal control problems, including SNN methods. Our method is completely determined by the formulation of DC which is a suitable DC–decomposition of the original optimal control problem. We present the algorithm in a function space setting in the spirit of [2], keeping in mind that there is an intermediary discretization procedure.

The DC–Algorithm is based on the fact that: if $\bar{u}$ is the solution of the primal problem ($P$) then $\partial H(\bar{u}) \subset \partial G(\bar{u})$ and conversely, if $u^*$ is the solution of the dual problem denoted by ($P^*$) we have the inclusion $\partial G^*(u^*) \subset \partial H^*(w^*)$, where $H^*$ and $G^*$ correspond to the dual functions of $H$ and $G$ respectively. In [2] there is an abstract framework for the DC algorithm in Banach spaces. Although, functions $G$ and $H$ do not satisfy all assumptions in [2], some of the results in [2] can be extended to our case with slightly modifications. In particular, if we define the function

$$
L(u, w) = G^*(w) - (w, u)_{L^2(\Omega)} + H(u).
$$

(40)

We can interpret optimality conditions for ($P_\gamma$) in terms of $L$. Indeed, if $\bar{w} \in \partial H(\bar{u}) \subset \partial G(\bar{u})$ then we have that $\bar{u} \in \partial G^*(\bar{w})$. This is equivalent to the following condition:

$$
\begin{align*}
L(\bar{u}, w) &\geq L(\bar{u}, \bar{w}), \\
L(u, \bar{w}) &\geq L(\bar{u}, \bar{w}),
\end{align*}
$$

(41)

(42)
for all $u$, and all $w$ in $L^2(\Omega)$. $(\bar{u}, \bar{w})$ is referred as $\partial$–critical point of $L$, see [2].

This symmetry means that DC–Algorithm alternates in computing approximations of the solutions for the primal and the dual problems as follows:

(43a) First chose: $w_k \in \partial H(u_k)$,

(43b) then chose: $u_k \in \partial G^*(w_k)$.

A more detailed discussion on the DC method can be found in [10] and [2]. In particular, in [2] the authors study the convergence properties for the DC algorithm in abstract spaces that covers our case with small changes.

Let us give a precise meaning to the numerical problems generated by (43). In view of the identity $\partial H(u_k) = \{w_k\}$, then, formula (19) implies that $w_k$ is given by

$$w_k = \begin{cases} 0, & \text{if } |u_k(x)| \leq \frac{1}{\gamma}, \\ \delta - \frac{1}{p} \left( |u_k(x)| + \frac{1-p}{\gamma} \right) \cdot \text{sign}(u_k(x)), & \text{otherwise}. \end{cases}$$

Moreover, according to Rockafellar [22] the subgradients can be computed as:

$$\partial G(y) = \text{argmax}_w \{ \langle y, w \rangle - G^*(w) \},$$

$$\partial G^*(w) = \text{argmax}_z \{ \langle w, z \rangle - G(z) \},$$

therefore, according to (46), $u_k$ can be obtained by solving the following optimal control problem

$$\min_{u_{k+1}} \frac{1}{2} \| Su_{k+1} + Sf - y_d \|_{L^2(\Omega)}^2 + \frac{\alpha}{2} \| u_{k+1} \|_{L^2(\Omega)}^2 + \delta \beta \| u_{k+1} \|_{L^1(\Omega)} - \int_{\Omega} w_k u_{k+1} \, dx.$$  

In case of the presence of box–constraints on the control, our formulation yields an box–constrained $L^1$ optimal control subproblem

$$\min_{u_k} \frac{1}{2} \| Su_k - y_d \|_{L^2(\Omega)}^2 + \frac{\alpha}{2} \| u_k \|_{L^2(\Omega)}^2 + \delta \beta \| u_k \|_{L^1(\Omega)} - \int_{\Omega} w_k u_k \, dx.$$

subject to:

$$u_k \in U_{ad}.$$

Remark 5. By the form of the DC splitting (DC) we replace problem (45) by the direct computation of $w_k$ from formula (44). In addition, observe that problem (47) is a convex $L^1$–sparse optimal control problem with penalization parameter $\delta \beta$, for which it is known to have a unique solution for $\alpha > 0$ c.f. [23]. The case of $\alpha = 0$ with box–constraints is also possible. Moreover, this problem can be solved numerically in an efficient way. For example, it can be solved by semi–smooth Newton methods proposed in [23] or, it can be solved in the framework of sparse programming problems in finite dimensions after its discretization.

In order to complete our algorithm, we now turn our attention to the following mechanism as stopping criterion. Looking at the gradient equation (23c) we could consider checking

$$\zeta_k = \frac{1}{\beta \delta} (w_k - \phi_k - \alpha u_k) \in \partial \| \cdot \|_{L^1(\Omega)}(u_k),$$
where $u_k$, $\phi_k$, $w_k$ represent the corresponding approximations of the optimal control, the adjoint state and the multipliers in the $k$–th iteration. In practice, a less sensitive stopping criterion gave us better results. This consists in checking the residual:

\[ \| \zeta_{k+1} - \zeta_k \| \leq \text{tol}, \]

where $\text{tol}$ is a prescribed tolerance. Notice that $\zeta_k$ contains information of the sign of the approximated solution $u_k$. Therefore, this condition is a analogous (but weaker) to the comparison of consecutive active sets, see for example [8]. Other stopping criteria can be considered as well, for example, a prescribed tolerance for the residual.

**Algorithm 1** DCA for problem $(P_\gamma)$

1: Initialize $u^0$.
2: while stopping criteria is false do
3: Compute $w_k$ given by (44)
4: Compute $u_{k+1}$ by solving problem (47) or (48) in case of control constraints.
5: $k \leftarrow k + 1$.
6: end while

4.1. **Advantages and disadvantages of DC–Algorithm.** Algorithm 1 theoretically is a first–order method, which provides a primal–dual updating procedure without need of the line–search step. This is an important feature in optimal control problems where the line–search step requires the evaluation of the cost function and its gradient in each iteration, requiring the computation of the state and adjoint equations, which usually are very expensive to solve numerically. Although in the proposed DC method there is no need of a line–search procedure, the methods to solve the subproblems are not PDE–free. In fact, the computational cost is concentrated in solving the subproblem which, in turn, might require a line–search procedure to solve the subproblem, depending on the method used for solving it. However, this is not an explicit feature of DCA.

In order to solve the subproblem (47), depending on our needs, one may apply either superlinear methods, like the Semi–smooth Newton Method (SNN), for solving the subproblem which are memory consuming or, descent second–methods, which solve smaller systems but usually they require line–search procedures adding its computational cost to the overall method.

Alternatively to DC–algorithm, semi–smooth Newton methods can be applied, in order to solve the optimality systems (23) and (38) directly. This leads to efficient super–linear methods c.f. [18]. However, we will require solving a coupled system involving the state $y$, the adjoint state $\phi$, and the multipliers $\zeta$, $\lambda_a$ and $\lambda_b$, resulting in a large system of equations, whose size depends on the discretization and the dimensionality of the domain. The possibility of using methods, which require only the computation of the state, to solve the subproblems (47) or (48) decreases the computational cost of solving linear systems. This is particularly useful when we deal with 3D problems, whose size escalate very quickly in fine meshes and require more memory resources. We summarize the numerical properties of the algorithm in the following table.

Whereas the primal dual algorithm proposed in [19] solves a large sparse system efficiently, the DC–Algorithm requires the solution of a sparse optimal control problem. In our setting we chose to use descent methods, intended specifically for $L^1$–sparse problems c.f. [9], which have the advantage of recovering the sparse components of the solution as null components. In contrast, PDA which computes sparsity only approximately close to 0.
We also observe that the DC–algorithm requires the tuning of more parameters, depending on the method used to solve the subproblems. This is a drawback when compared with PDA, which only requires choosing one regularization parameter $\varepsilon$.

Finally, we mention that both algorithms can be combined. For example, after obtaining a solution using PDA we can recover the sparse components by refining the solution as the input for DCA.

5. Implementation aspects

5.1. Approximation. For simplicity, the approximation of problems $[P]$ and $[PC]$ is done by the finite–difference scheme, although other discretization methods might be applied such as finite elements. Uniform meshes are considered in the domain $\Omega$ with $N$ internal nodes. The associated mesh parameter is given by $h = \frac{1}{N+1}$. Then, the state equation (3) is solved numerically with the finite difference method, while the approximation of the integrals is computed using the following mid–point rule:

\begin{equation}
\int_a^b \int_c^d u(x,y)dydx \approx \frac{1}{4}h^2 \left\{ u(a,c) + u(b,c) + u(a,d) + u(b,d) \right. \\
+ 2 \sum_{i=1}^{n-2} u(x_i,c) + 2 \sum_{i=1}^{n-2} u(x_i,d) + 2 \sum_{i=1}^{n-2} u(a,y_i) \\
+ 2 \sum_{i=1}^{n-2} u(b,y_i) + 4 \sum_{i=1}^{n-2} \sum_{j=1}^{n-2} u(x_i,y_i) \right\}.
\end{equation}

Using this approximation, and reshaping the matrix $(u(x_i,y_j))_{i,j=1,\ldots,N}$ as a vector $u \in \mathbb{R}^{N^2}$ the $L^1$–norm is approximated by

\begin{equation}
\|u\|_1 \approx \sum_{i=1}^{N^2} c_i |u_i|,
\end{equation}

where the $c_i$'s are the corresponding coefficients given by (51).

5.2. Auxiliar $L^1$–sparse optimal control problems. DC–algorithm [1] has a simple structure. However, they require to solve auxiliar $L^1$–norm optimal control subproblems (47) and (48) respectively. Clearly, the efficiency of the proposed algorithms strongly depends on the numerical methods applied for solving (47) and (48). As mentioned earlier, the numerical solution of the $L^1$–norm optimal control problems can be done by semi–smooth Newton methods as in [23]. However, semi–smooth Newton methods do not guarantee a reduction in the cost function in each iteration. On the other hand, many methods for solving $L^1$–norm functionals are known to be of first–order since no regularization procedure is considered. Although, several methods in finite dimensions can be
used, we take advantage of OESOM method which is suitable for this class of optimal control problems, see [9] for details and numerical evidence of the method.

In contrast with several descent methods, the application of OESOM algorithm proposed in [9] is straightforward. Indeed, we only need to provide the cost function and the corresponding gradient which involves the computing of the adjoint state. The last one can be evaluated by means of the adjoint state (23b), (37b). Second order information of the smooth can be considered using Hessians or its approximation by means of BFGS or LBFGS methods. In addition, approximated second order information of non differentiable term is calculated by the built-in enriched second order information constructed by the OESOM algorithm using weak derivatives of the $L^1$-norm, see [9] for the details.

6. Numerical Evidence

In order to investigate the numerical performance of the proposed DC–algorithm in Section 4 we have implemented Algorithm 1 using MATLAB. The associated sparse $L^1$ subproblem was solved using the OESOM algorithm [9] by extending it to the box–constrained case with an additional projection step on the admissible control set. The OESOM algorithm is a second–order method for solving $\ell_1$–norm penalized optimization problems which includes second order information hidden in the structure of the $\ell_1$–norm, therefore, is an efficient method to solve the $L^1$–norm subproblem from the DC–algorithm.

As illustrative examples, we consider the following tests defined on the unit square domain $\Omega = (0,1) \times (0,1)$. The corresponding PDEs were approximated using finite differences.

Example 1. We consider problem $(P)$ for $A = -\Delta$ and $y_d = e^{-\cos(2\pi xy)^2/0.1}$.

Performance of a single run. We first solve this example fixing the values of $\alpha = 1/4$ and $\beta = 7/10$. Algorithm 1 gives an approximated solution after 19 iterations stopping when $\|\zeta_{k+1} - \zeta_k\| < tol = 1e - 5$. The table and graphics below, show the performance and behavior of DCA. We observe in Figure 2 with logarithmic scale in the $x$ axis, the decreasing behavior of the objective function is more intensive in the first iterations. We also show the decreasing of the distance of consecutive approximated multipliers in the logarithmic scale in the $y$ axis.

![Figure 2](image)

**Figure 2.** Cost function and residual of $\zeta$ at $\beta = 0.004$

*Figure 2 (right)* depicts the evolution of stopping criteria, which is more erratic with a decreasing tendency. In each iteration new sparse components appear then, when comparing consecutive multipliers, they may differ from 0 to 1 in those components, causing oscillations on their difference. We also realize in Table 2 that the number of sparse components of the approximated solution is increasing at every iterate.
Table 2. Performance data for DCA for Example 1

Varying the regularization parameter $\gamma$. According to our theory, it is expected that if $\gamma \to \infty$ the solution $\bar{u}_\gamma \to \bar{u}$. Here, we solve Example 1 for increasing values of $\gamma$. The numerical evidence of this convergence behavior is reflected in Table 3 where we observe optimal cost converges to a fixed value, whereas sparsity also stabilizes at 1525 null components of the solution.

Table 3. Numerical convergence for increasing values of $\gamma$.

Varying the regularization parameter $\beta$. Now we experiment with different values of $\beta$, which determines the sparsity-inducing term $\Upsilon$. Table 4 shows that larger values of $\beta$ result in sparser solutions until the solution vanishes, which illustrates Lemma 1. As expected, it can also be observed that the optimal cost increases accordingly to the sparsity of the solution, reflected in smaller supports of the controls.

Varying the exponent $p$. We finish this example with the variation of the fractional exponent $1/p$ which also plays a role in the sparsity of the solution. In fact, $p$ determines
Table 4. Solutions become sparser as $\beta$ increases.

| $\beta$ | Optimal Cost | Sparse components | DCA Iterations |
|---------|--------------|-------------------|----------------|
| 0.0002  | 229.1145     | 1034              | 25             |
| 0.0005  | 229.259      | 1729              | 30             |
| 0.0010  | 229.4327     | 2528              | 37             |
| 0.0015  | 229.5503     | 3004              | 30             |
| 0.0020  | 229.6252     | 3359              | 31             |
| 0.0025  | 229.6676     | 3631              | 30             |
| 0.0030  | 229.6849     | 3843              | 37             |

Figure 3. Optimal control and its support for $\beta = 0.0002$.

Figure 4. Optimal control and its support for $\beta = 0.001$.

how expensive is a sparse control. It is known that for larger values of $p$ the sparsity term tends to produce a volume constraint induced by the Donoho’s counting norm c.f.\cite{19}. However, the increment of $p$ does not necessarily increase sparsity in the solution as we can see in Table 5.
Figure 5. Optimal control and its support for $\beta = 0.002$.

Figure 6. Optimal control and its support for $\beta = 0.003$.

Table 5. Influence of the power parameter $p$ in the sparsity of the solution.

| $p$ | Optimal Cost | Sparse components | DCA Iterations |
|-----|--------------|-------------------|----------------|
| 1   | 229.2028     | 789               | 4              |
| 1.2 | 229.3232     | 1860              | 18             |
| 1.5 | 229.4736     | 2778              | 32             |
| 2   | 229.6256     | 3355              | 27             |
| 4   | 229.8814     | 3667              | 26             |
| 8   | 230.3485     | 3441              | 23             |
| 10  | 230.5699     | 3323              | 28             |
| 20  | 231.4921     | 2846              | 23             |
Example 2. In this example, we compare DC–Algorithm with the primal–dual method proposed in [19] (See eq. (5.7), pg. 1273 for problem (P_{n,e})) developed to solve optimal control problems involving $L^q$-penalizations with $q \in (0,1)$. Here, we consider a $L^2$ penalization on the gradient of the control. Therefore, the control space is restricted to a subset of $H^1_0(\Omega)$. Although, this penalization is beyond our theory, it can be considered with straightforward modifications. The problem reads

$$\min_{(y,u)} \frac{1}{2} \|y - y_d\|^2_2(\Omega) + \frac{1}{2} \|\nabla u\|^2_2(\Omega) + \beta \gamma (u)$$

(E2)

subject to

$$\begin{align*}
-\Delta y &= u, \quad \text{in } \Omega, \\
y &= 0, \quad \text{on } \Gamma,
\end{align*}$$

In the framework of [19], we choose the quantities $B = I$, $E = -\Delta$, $K = E^{-1}$, $g = 0$, $f = y_d$ and $Y = L^2(\Omega)$. Therefore, the numerical scheme (5.7) in [19] consist in the sequence of equations of the form:

$$-\Delta u_{k+1} + K^*Ku_{k+1} + \frac{\beta/p}{\max(\varepsilon^{2-1/p}, |u_k|^{2-1/p})}u_{k+1} = K^*y_d, \quad k = 0,1,2,\ldots$$

(53)

The operator $K$ and $K^*$ involve the inverse of the differential operator (corresponding to the laplacian, in this example). However, it is an uncommon situation having an explicit representation of $K$ and $K^*$, rather we have to solve the associated PDE. Therefore, we introduce the state $y_{k+1}$ and the adjoint state $\phi_{k+1}$. Then, equation (53) is reformulated as the following iterative system:

$$\begin{pmatrix}
\alpha E + \frac{\beta/p}{\max(\varepsilon^{2-1/p}, |u_k|^{2-1/p})} & 0 & I \\
0 & E & -I \\
-I & 0 & E
\end{pmatrix}
\begin{pmatrix}
\begin{pmatrix} u_{k+1} \\ \phi_{k+1} \\ y_{k+1} \end{pmatrix}
\end{pmatrix} =
\begin{pmatrix}
0 \\
y_d \\
0
\end{pmatrix}.$$  

(54)

In order to compare DC-Algorithm (DCA) with the Primal Dual based Algorithm (54) (which we will refer as PD-Algorithm, PDA for short) we observe their performance at different values of the regularization parameters with both methods starting from the same initial point $u_0$. There is not direct relation between the regularization parameters $\varepsilon$ of PD-Algorithm and $\gamma$ used in DC-Algorithm. Therefore, we chose regularization parameters for each regularizer such that the function $|t|^{1/p}$ with approximately the same error, i.e. the regularization error satisfies: $R_e = \| |t|^{1/p} - t_r \|_\infty \approx \text{tol}$, where tol is a tolerance and $t_r$ denotes the regularization.

Moreover, since both algorithms have different stopping rules we observe the cost value after 100 iterations, to guarantee that both algorithms are close enough to the solution. The results are summarized in Table 6. In our experiments we found a similar performance of both algorithms since they results have close approximated optimal costs. After 100 iterations we observe that PDA or DCA can reach the minimum cost, depending on the regularization parameters.

| Reg. | Re | Cost ($\beta = 0.005$) | Cost ($\beta = 0.01$) | Cost ($\beta = 0.2$) |
|------|----|------------------------|------------------------|------------------------|
| PD-Algorithm | $\varepsilon = 0.0001$ | 0.00750 | 6.10089 | 7.2573 | 8.30983 |
| DC-Algorithm | $\gamma = 500$ | 0.00746 | 6.10053 | 7.1810 | 8.19439 |
| | $\gamma = 300$ | 0.01298 | 6.10050 | 7.1915 | 8.19786 |

Table 6. Comparison with primal-dual algorithm after 100 iterations.
Figure 7 shows sparse components of the solution computed by PDA and DCA methods respectively. It can be observed that PD–Algorithm computes sparse components approximately $0 \approx 10^{-4}$ while DC–Algorithm is able to recover zero sparse components as expected from the theory.

**Example 3.** This example consists in imposing box–constraints on Example 1. We keep the same parameters as in Example 1. Therefore, we require in addition that

$$u \in U_{ad} = \{u \in L^2(\Omega) : 0 \leq u \leq 0.035\}.$$ 

Similar results are observed in this case as depicted in Figure 8. The structure of the sparsity and the support of the optimal control is similar but in this case the optimal control is also active on the prescribed bounds as observed in Figure 9.
Our final experiment is out of scope of this paper since our theory does not consider the case $\alpha = 0$. However, the method is still useful to this case and further analysis is required. Our problem consists in a box–constrained optimal control problem with $L^q$–term only ($\alpha = 0$). Here the desired state is $y_d(x_1, x_2) = \sin(2\pi x_1) \sin(2\pi x_2)$ and the set of admissible controls is given by

$$u \in U_{ad} = \{ u \in L^2(\Omega) : -0.035 \leq u \leq 0.035 \}.$$ 

In this case we observe (c.f. Figure 10) a typical bang–bang optimal control shape.

Figure 10. Box–constrained optimal control and its support for $\alpha = 0$. 
References

[1] Antonio Ambrosetti and Giovanni Prodi. *A primer of nonlinear analysis*. Number 34. Cambridge University Press, 1995.

[2] Giles Auchmuty. Duality algorithms for nonconvex variational principles. *Numerical Functional Analysis and Optimization*, pages 863–874. Volume 10, 9-10 Taylor & Francis, 1989.

[3] Eduardo Casas, Christian Clason, and Karl Kunisch. Parabolic control problems in measure spaces with sparse solutions. *SIAM Journal on Control and Optimization*, 51(1):28–63, 2013.

[4] Eduardo Casas and Karl Kunisch. Parabolic control problems in space-time measure spaces. *ESAIM: Control, Optimisation and Calculus of Variations*, 22(2):355–370, 2016.

[5] Eduardo Casas. A review on sparse solutions in optimal control of partial differential equations. *SeMA Journal*, volume 74(3):319–344, 2017.

[6] Philippe G Ciarlet. *Linear and nonlinear functional analysis with applications*, volume 130. Siam, 2013.

[7] Gianni Dal Maso. *An introduction to Γ-convergence*. Springer Science & Business Media, volume 8, 2012.

[8] Juan Carlos De los Reyes. Theory of PDE-constrained optimization. In *Numerical PDE-Constrained Optimization*, pages 25–41. Springer International Publishing, 2015.

[9] Juan Carlos De Los Reyes, Estefanía Loayza, and Pedro Merino. Second-order orthant–based methods with enriched hessian information for sparse ℓ1-optimization. *Computational Optimization and Applications*, 67(2):225–258, 2017.

[10] Tao Pham Dinh and Hoai An Le Thi. Recent advances in dc programming and DCA. In *Transactions on Computational Intelligence XIII*, pages 1–37. Springer, 2014.

[11] Fabián Flores-Bazán and W Oettli. Simplified optimality conditions for minimizing the difference of vector-valued functions. *Journal of Optimization Theory and Applications*, 108(3):571–586, 2001.

[12] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*, volume 1. Birkhäuser Basel, 2013.

[13] Michael Hintermüler and Tao Wu. Nonconvex TVp–models in image restoration: analysis and a trust-region regularization–based superlinearly convergent solver. *SIAM Journal on Imaging Sciences*, 6(3):1385–1415, 2013.

[14] Michael Hinzke, René Pinnau, Michael Ulbrich, and Stefan Ulbrich. *Optimization with PDE constraints*, volume 23. Springer Science & Business Media, 2008.

[15] Jean-Baptiste Hiriart-Urruty. From convex optimization to nonconvex optimization. Necessary and sufficient conditions for global optimality. In *Nonsmooth optimization and related topics*, pages 219–239. Springer, 1989.

[16] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Fundamentals of convex analysis*. Springer Science & Business Media, 2012.

[17] Aleksandr Davidovič Ioffe, Vladimir Mikhailovič Tihomirov, and Bernd Luderer. *Theorie der ExtremalAufgaben*. VEB Deutscher Verlag der Wissenschaften, 1979.

[18] Kazufumi Ito and Karl Kunisch. Lagrange multiplier approach to variational problems and applications. *SIAM Advances in Design and Control*, 52(2):1251–1275, 2014.

[19] Kazufumi Ito and Karl Kunisch. Optimal control with $L^p(\Omega)$, $p \in [0,1)$, control cost. *SIAM Journal on Control and Optimization*, Philadelphia, 2008.

[20] Johannes Jahn. *Introduction to the Theory of Nonlinear Optimization*. Springer-Verlag Berlin Heidelberg, 3 edition, 2007.

[21] Ronny Ramlau and Clemens A Zarzer. On the minimization of a Tikhonov functional with a non-convex sparsity constraint. *Electronic Transactions on Numerical Analysis*, 39:476–507, 2012.

[22] Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.

[23] G. Stadler. Elliptic optimal control problems with $L^2$-control cost and applications for the placement of control devices. *Comput. Optim. Appl.*, 44(2):159–181, 2009.

1Research Center of Mathematical Modeling (MODEMAT) and Department of Mathematics, Escuela Politécnica Nacional, Quito, Ecuador