*IMA Journal of Numerical Analysis* (2018) **38**, 852–883
doi: 10.1093/imanum/drx018
Advance Access publication on May 29, 2017

# Some applications of weighted norm inequalities to the error analysis of PDE-constrained optimization problems

HARBIR ANTIL
*Department of Mathematical Sciences, George Mason University, Fairfax, VA 22030, USA*
hantil@gmu.edu

ENRIQUE OTÁROLA
*Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile*
enrique.otarola@usm.cl

AND

ABNER J. SALGADO*
*Department of Mathematics, University of Tennessee, Knoxville, TN 37996, USA*
*Corresponding author: asalgad1@utk.edu

[Received on 3 May 2016; revised on 21 February 2017]

The purpose of this work is to illustrate how the theory of Muckenhoupt weights, Muckenhoupt-weighted Sobolev spaces and the corresponding weighted norm inequalities can be used in the analysis and discretization of partial-differential-equation-constrained optimization problems. We consider a linear quadratic constrained optimization problem where the state solves a nonuniformly elliptic equation, a problem where the cost involves pointwise observations of the state and one where the state has singular sources, e.g., point masses. For all the three examples, we propose and analyse numerical schemes and provide error estimates in two and three dimensions. While some of these problems might have been considered before in the literature, our approach allows for a simpler, Hilbert-space-based analysis and discretization and further generalizations.

*Keywords*: PDE-constrained optimization; Muckenhoupt weights; weighted Sobolev spaces; finite elements; polynomial interpolation in weighted spaces; nonuniform ellipticity; point observations; singular sources.

## 1. Introduction

The purpose of this work is to show how the theory of Muckenhoupt weights, Muckenhoupt-weighted Sobolev spaces and weighted norm inequalities can be applied to analyse partial-differential-equation(PDE)-constrained optimization problems and their discretizations. These tools have already been shown to be essential in the analysis and discretization of problems constrained by equations involving fractional derivatives both in space and in time (Antil & Otárola, 2015; Antil *et al.*, 2016), and here we extend their use to a new class of problems.

We consider three illustrative examples. While some of them have been considered before, the techniques that we present in this study are new and we believe they provide simpler arguments and allow for further generalizations. To describe them, let $\Omega$ be an open and bounded polytopal domain of $\mathbb{R}^n$ ($n \in \{2, 3\}$), with Lipschitz boundary $\partial\Omega$. We will be dealing with the following problems:

- *Optimization with nonuniformly elliptic equations.* Let $\omega$ be a weight, that is, an almost everywhere positive and locally integrable function and $y_d \in L^2(\omega, \Omega)$. Given a regularization parameter $\lambda > 0$, we define the cost functional

$$J_{\mathcal{A}}(y,u) = \frac{1}{2}\|y - y_d\|^2_{L^2(\omega,\Omega)} + \frac{\lambda}{2}\|u\|^2_{L^2(\omega^{-1},\Omega)}. \tag{1.1}$$

We are then interested in finding min $J_{\mathcal{A}}$ subject to the *nonuniformly elliptic problem*

$$-\text{div}(\mathcal{A}\nabla y) = u \text{ in } \Omega, \qquad y = 0 \text{ on } \partial\Omega, \tag{1.2}$$

and the control constraints

$$u \in \mathbb{U}_{\mathcal{A}}, \tag{1.3}$$

where $\mathbb{U}_{\mathcal{A}}$ is a nonempty, closed and convex subset of $L^2(\omega^{-1}, \Omega)$. The main source of difficulty and originality here is that the matrix $\mathcal{A}$ is not uniformly elliptic, but rather satisfies

$$\omega(x)|\xi|^2 \lesssim \xi^{\mathrm{T}} \cdot \mathcal{A}(x) \cdot \xi \lesssim \omega(x)|\xi|^2 \quad \forall\, \xi \in \mathbb{R}^n, \tag{1.4}$$

for almost every $x \in \Omega$. Since we allow the weight to vanish or blow up, this nonstandard ellipticity condition must be treated with the right functional setting.

Problems such as (1.2) arise when applying the so-called Caffarelli–Silvestre extension for fractional diffusion (Caffarelli & Silvestre, 2007; Antil & Otárola, 2015; Nochetto *et al.*, 2015, 2016; Antil *et al.*, 2016), when dealing with boundary controllability of parabolic and hyperbolic degenerate equations (Cannarsa *et al.*, 2008; Du, 2014; Gueye, 2014) and in the numerical approximation of elliptic problems involving measures (Agnelli *et al.*, 2014; Nochetto *et al.*, 2016). In addition, invoking Rubio de Francia's extrapolation theorem (Duoandikoetxea, 2001, Theorem 7.8), one can argue that this is a quite general PDE-constrained optimization problem with an elliptic equation as state constraint, since *there is no* $L^p$*, only* $L^2$ *with weights.*

- *Optimization with point observations.* Let $\emptyset \neq \mathcal{Z} \subset \Omega$ with $\#\mathcal{Z} < \infty$. Given a set of prescribed values $\{y_z\}_{z\in\mathcal{Z}}$, a regularization parameter $\lambda > 0$, and the cost functional

$$J_{\mathcal{Z}}(y,u) = \frac{1}{2}\sum_{z\in\mathcal{Z}} |y(z) - y_z|^2 + \frac{\lambda}{2}\|u\|^2_{L^2(\Omega)}, \tag{1.5}$$

the problem under consideration reads as follows: find min $J_{\mathcal{Z}}$ subject to

$$-\Delta y = u \text{ in } \Omega, \qquad y = 0 \text{ on } \partial\Omega, \tag{1.6}$$

and the control constraints

$$u \in \mathbb{U}_{\mathcal{Z}}, \tag{1.7}$$

where $\mathbb{U}_{\mathcal{Z}}$ is a nonempty, closed and convex subset of $L^2(\Omega)$. In contrast to standard elliptic PDE-constrained optimization problems, the cost functional (1.5) involves point evaluations of the state.

We must immediately comment that since $\partial\Omega$ is Lipschitz and $f \in L^2(\Omega)$ then there exists $r > n$ such that $y \in W^{1,r}(\Omega)$ (Jerison & Kenig, 1995, Theorem 0.5; see also Jerison & Kenig, 1981; Grisvard, 1985; Dauge, 1992; Savaré, 1998; Maz'ya & Rossmann, 2010). This, on the basis of a Sobolev embedding result, implies that $y \in C(\bar{\Omega})$ and thus that the point evaluations of the state $y$ in (1.5) are well defined, the latter leading to a subtle formulation of the adjoint problem (see Section 4 for details).

Problem (1.5)–(1.7) finds relevance in numerous applications where the observations are carried out at specific locations: for instance, in the so-called calibration problem with American options (Achdou, 2005), in the optimal control of selective cooling of steel (Unger & Tröltzsch, 2001), in the active control of sound (Nelson & Elliott, 1992; Bermúdez *et al.*, 2004) and in the active control of vibrations (Fuller *et al.*, 1996; Hernández & Otárola, 2009). See also Rannacher & Vexler (2005), Hintermüller & Laurain (2008), Gong *et al.* (2014a), Brett *et al.* (2015), Brett *et al.* (2016) for other applications. The point observation terms in the cost (1.5) tend to enforce the state $y$ to have the fixed value $y_z$ at the point $z$. Consequently, (1.5)–(1.7) can be understood as a penalty version of a PDE-constrained optimization problem where the state is constrained at a collection of points. We refer the reader to Brett *et al.* (2016, Section 3.1) for a precise description of this connection and to Leykekhman *et al.* (2013) for the analysis and discretization of an optimal control problem with state constraints at a finite number of points.

Despite its practical importance, to the best of our knowledge, there are only two references where the approximation of (1.5)–(1.7) is addressed: Chang *et al.* (2015) and Brett *et al.* (2016). In both works the key observation, and main source of difficulty, is that the adjoint state for this problem is only in $W_0^{1,r}(\Omega)$ with $r \in (\frac{2n}{n+2}, \frac{n}{n-1})$. With this functional setting, the authors of Brett *et al.* (2016) propose a fully discrete scheme that discretizes the control explicitly using *piecewise linear elements*. For $n = 2$, the authors obtain an $\mathcal{O}(h)$ rate of convergence for the optimal control in the $L^2$ norm, provided the control and the state are discretized using meshes of size $\mathcal{O}(h^2)$ and $\mathcal{O}(h)$, respectively (see Brett *et al.*, 2016, Theorem 5.1). This condition immediately poses two challenges for implementation. First, it requires keeping track of the state and control on different meshes. Second, some sort of interpolation and projection between these meshes needs to be realized. In addition, the number of unknowns for the control is significantly higher, thus leading to a slow optimization solver. The authors of Brett *et al.* (2016) were unable to extend these results to $n = 3$. Using the so-called variational discretization approach (Hinze, 2005), the control is implicitly discretized, and the authors were able to prove that the control converges with rates $\mathcal{O}(h)$ for $n = 2$ and $\mathcal{O}(h^{1/2-\epsilon})$ for $n = 3$. In a similar fashion, the authors of Chang *et al.* (2015) use the variational discretization concept to obtain an implicit discretization of the control and deduce rates of convergence of $\mathcal{O}(h)$ and $\mathcal{O}(h^{1/2})$ for $n = 2$ and $n = 3$, respectively. A residual-type *a posteriori* error estimator is introduced, and its reliability is proven. However, there is no analysis of the efficiency of the estimator.

In Section 4, we introduce a fully discrete scheme where we discretize the control with piecewise constants; this leads to a smaller number of degrees of freedom for the control in comparison with the approach of Brett *et al.* (2016). We circumvent the difficulties associated with the adjoint state by working in a weighted $H^1$ space and prove the following rates of convergence for the optimal control: $\mathcal{O}(h|\log h|)$ for $n = 2$ and $\mathcal{O}(h^{\frac{1}{2}}|\log h|)$ for $n = 3$. In addition, we provide pointwise error estimates for the approximation of the state: $\mathcal{O}(h|\log h|)$ for $n = 2$ and $\mathcal{O}(h^{1/2}|\log h|)$ for $n = 3$.

- *Optimization with singular sources.* Let $\mathcal{D} \subset \Omega$ be linearly ordered and with cardinality $l = \#\mathcal{D} < \infty$. Given a desired state $y_d \in L^2(\Omega)$ and a regularization parameter $\lambda > 0$, we define the cost functional

$$J_\delta(y,\mathbf{u}) = \frac{1}{2}\|y-y_d\|^2_{L^2(\Omega)} + \frac{\lambda}{2}\|\mathbf{u}\|^2_{\mathbb{R}^l}. \tag{1.8}$$

We shall be concerned with the following problem: find min $J_\delta$ subject to

$$-\Delta y = \sum_{z\in\mathcal{D}} u_z\delta_z \text{ in } \Omega, \qquad y=0 \text{ on } \partial\Omega, \tag{1.9}$$

where $\delta_z$ is the Dirac delta at the point $z$ and

$$\mathbf{u} = \{u_z\}_{z\in\mathcal{D}} \in \mathbb{U}_\delta, \tag{1.10}$$

where $\mathbb{U}_\delta \subset \mathbb{R}^l$ with $\mathbb{U}_\delta$, again, nonempty, closed and convex. Notice that since for $n>1$, $\delta_z \notin H^{-1}(\Omega)$, the solution $y$ to (1.9) does not belong to $H^1(\Omega)$. Consequently, the analysis of the finite element method applied to such a problem is not standard (Scott, 7374; Casas, 1985; Nochetto *et al.*, 2016). We rely on the weighted Sobolev space setting described and analysed in Nochetto *et al.* (2016, Section 7.2).

The state (1.9), in a sense, is dual to the adjoint equation for (1.5)–(1.6): it is an elliptic equation that has Dirac deltas on the right-hand side. The optimization problem (1.8)–(1.9) is of relevance in applications where one can specify a control at finitely many prespecified points. For instance, some works (Nelson & Elliott, 1992; Bermúdez *et al.*, 2004) discuss applications within the context of the active control of sound (Fuller *et al.*, 1996; Hernández & Otárola, 2009; Hernández *et al.*, 2010) and in the active control of vibrations (see also Leykekhman & Vexler, 2013; Fornasier *et al.*, 2014; Gong *et al.*, 2014a).

An analysis of problem (1.8)–(1.10) is presented in Gong *et al.* (2014b), where the authors use the variational discretization concept to derive error estimates. They show that the control converges with a rate of $\mathcal{O}(h)$ and $\mathcal{O}(h^{1/2})$ in two and three dimensions, respectively. Their technique is based on the fact that the state belongs to $W^{1,r}_0(\Omega)$ with $r \in (\frac{2n}{n+2}, \frac{n}{n-1})$. In addition, under the assumption that $y_d \in L^\infty(\Omega)$ they improve their results and obtain, up to logarithmic factors, rates of $\mathcal{O}(h^2)$ and $\mathcal{O}(h)$. Finally, we mention that Casas *et al.* (2012) and Pieper & Vexler (2013) study a PDE-constrained optimization problem without control constraints, but where the control is a regular Borel measure.

In Section 5, we present a fully discrete scheme for which we provide rates of convergence for the optimal control: $\mathcal{O}(h^{2-\epsilon})$ in two dimensions and $\mathcal{O}(h^{1-\epsilon})$ in three dimensions, where $\epsilon>0$. We also present rates of convergence for the approximation error in the state variable.

Before we embark on further discussions, we must remark that while the introduction of a weight as a technical instrument does not seem to be completely new, the techniques that we use and the range of problems that we can tackle is. For instance, for integro-differential equations where the kernel $g$ is weakly singular, the authors of Burns & Ito (1995) study the well-posedness of the problem in the weighted $L^2(g,(-r,0))$ space. Numerical approximations for this problem with the same functional setting were considered in Ito & Turi (1991), where convergence is shown, but no rates are obtained. These ideas were extended to neutral delay-differential equations in Fabiano & Turi (2003) and Fabiano (2013), where a weight is introduced in order to renorm the state space and obtain dissipativity of the underlying operator. In all these works, however, the weight is essentially assumed to be smooth and monotone, except at the origin where it has an integrable singularity (Ito & Turi, 1991; Burns & Ito, 1995) or at a finite number of points where it is allowed to have jump discontinuities (Fabiano & Turi, 2003; Fabiano, 2013). All these properties are used to obtain the aforementioned results. In contrast, our approach hinges *only* on the fact that the introduced weights belong to the Muckenhoupt class $A_2$ (see Definition 2.1 below) and the pertinent facts from real and harmonic analysis and approximation theory that follow from this definition. Additionally, we obtain convergence rates for the optimal control variable that are, in terms of approximation, optimal for problem (1.1)–(1.3), nearly optimal in two dimensions and suboptimal in three dimensions for (1.5)–(1.7) and suboptimal for problem (1.8)–(1.10). Finally, we must point out that the class of problems we study is quite different from those considered in the references given above.

Our presentation will be organized as follows. Notation and general considerations will be introduced in Section 2. Section 3 presents the analysis and discretization of problem (1.1)–(1.3). Problem (1.5)–(1.7) is studied in Section 4. The analysis of problem (1.8)–(1.10) is presented in Section 5. Finally, in Section 6, we illustrate our theoretical developments with a series of numerical examples.

## 2. Notation and preliminaries

Let us fix notation and the setting in which we will operate. In what follows, $\Omega$ is a convex, open and bounded domain of $\mathbb{R}^n$ ($n \geq 1$) with polytopal boundary. The handling of curved boundaries is somewhat standard but leads to additional technicalities that will only obscure the main ideas we are trying to advance. By $A \lesssim B$ we mean that there is a nonessential constant $c$ such that $A \leq cB$. The value of this constant might change at each occurrence.

### 2.1 *Weights and weighted spaces*

Throughout our discussion we call a *weight* a function $\omega \in L^1_{\text{loc}}(\mathbb{R}^n)$, such that $\omega(x) > 0$ for a.e. $x \in \mathbb{R}^n$. In particular, we are interested in the so-called *Muckenhoupt* weights (Turesson, 2000; Duoandikoetxea, 2001).

Definition 2.1 (Muckenhoupt class.) Let $r \in (1, \infty)$ and $\omega$ be a weight. We say that $\omega \in A_r$ if

$$C_{r,\omega} := \sup_B \left( \fint_B \omega(x)\,\mathrm{d}x \right) \left( \fint_B \omega^{1/(1-r)}(x)\,\mathrm{d}x \right)^{r-1} < \infty,$$

where the supremum is taken over all balls $B \subset \mathbb{R}^n$.

Every weight induces a measure $\omega\,\mathrm{d}x$. For a measurable $E \subset \mathbb{R}^n$, we define

$$\omega(E) = \int_E \omega(x)\,\mathrm{d}x, \qquad \omega^{-1}(E) = \int_E \omega^{-1}(x)\,\mathrm{d}x. \tag{2.1}$$

From the fact that $\omega \in A_r$, many fundamental consequences for analysis follow. For instance, the induced measure $\omega\,\mathrm{d}x$ is not only doubling but also *strong doubling* (cf. Nochetto *et al.*, 2016, Proposition 2.2). We introduce the *weighted Lebesgue spaces*

$$L^r(\omega, \Omega) = \left\{ v \in L^0(\Omega) : \int_\Omega |v(x)|^r \omega(x)\,\mathrm{d}x < \infty \right\}$$

and note that Nochetto *et al.* (2016, Proposition 2.3) shows that their elements are distributions; therefore we can define *weighted Sobolev spaces*

$$W^{k,r}(\omega,\Omega)=\{v\in L^r(\omega,\Omega): D^\kappa v\in L^r(\omega,\Omega)\ \forall\kappa:|\kappa|\le k\},$$

which are complete and separable, and smooth functions are dense in them (cf. Turesson, 2000, Proposition 2.1.2, Corollary 2.1.6). We define $H^1(\omega,\Omega)=W^{1,2}(\omega,\Omega)$.

We define $W_0^{k,r}(\omega,\Omega)$ as the closure of $C_0^\infty(\Omega)$ in $W^{k,r}(\omega,\Omega)$ and set $H_0^1(\omega,\Omega)=W_0^{1,2}(\omega,\Omega)$. On these spaces, the following *Poincaré* inequality holds:

$$\|v\|_{L^r(\omega,\Omega)}\lesssim\|\nabla v\|_{L^r(\omega,\Omega)}\quad\forall v\in W_0^{1,r}(\omega,\Omega),\tag{2.2}$$

where the hidden constant is independent of $v$, depends on the diameter of $\Omega$ and depends on $\omega$ only through $C_{r,\omega}$.

The literature on the theory of Muckenhoupt-weighted spaces is rather vast, so we only refer the reader to Turesson (2000), Duoandikoetxea (2001) and Nochetto *et al.* (2016) for further results.

### 2.2 *Finite element approximation of weighted spaces*

Since the spaces $W^{1,r}(\omega,\Omega)$ are separable for $\omega\in A_r$ $(r>1)$, and smooth functions are dense, it is possible to develop a complete approximation theory using functions that are piecewise polynomial. This is essential, for instance, to analyse the numerical approximation of (1.2) with finite element techniques. Let us then recall the main results from Nochetto *et al.* (2016) concerning this scenario.

Let $\mathscr{T}=\{T\}$ be a conforming triangulation (into simplices or $n$-rectangles) of $\Omega$. We denote by $\mathbb{T}=\{\mathscr{T}\}$ a family of triangulations, which for simplicity we assume quasiuniform. The mesh size of $\mathscr{T}\in\mathbb{T}$ is denoted by $h_{\mathscr{T}}$. Given $\mathscr{T}\in\mathbb{T}$, we define the finite element space

$$\mathbb{V}(\mathscr{T})=\left\{v_{\mathscr{T}}\in C^0(\bar\Omega): v_{\mathscr{T}|T}\in\mathcal{P}(T),\ v_{\mathscr{T}|\partial\Omega}=0\right\},\tag{2.3}$$

where, if $T$ is a simplex, $\mathcal{P}(T)=\mathbb{P}_1(T)$—the space of polynomials of degree at most 1. In the case that $T$ is an $n$-rectangle, $\mathcal{P}(T)=\mathbb{Q}_1(T)$— the space of polynomials of degree at most 1 in each variable. Notice that, by construction, $\mathbb{V}(\mathscr{T})\subset W_0^{1,\infty}(\Omega)\subset W_0^{1,r}(\omega,\Omega)$ for any $r\in(1,\infty)$ and $\omega\in A_r$.

The results of Nochetto *et al.* (2016) show that there exists a quasi-interpolation operator $\Pi_{\mathscr{T}}: L^1(\Omega)\to\mathbb{V}(\mathscr{T})$, which is based on local averages over stars and thus is well defined for functions in $L^1(\Omega)$. This operator satisfies the following stability and approximation properties:

$$\begin{aligned}
\|\Pi_{\mathscr{T}}v\|_{L^r(\omega,\Omega)}&\lesssim\|v\|_{L^r(\omega,\Omega)} &&\forall v\in L^r(\omega,\Omega),\\
\|v-\Pi_{\mathscr{T}}v\|_{L^r(\omega,\Omega)}&\lesssim h_{\mathscr{T}}\|v\|_{W^{1,r}(\omega,\Omega)} &&\forall v\in W^{1,r}(\omega,\Omega),\\
\|\Pi_{\mathscr{T}}v\|_{W^{1,r}(\omega,\Omega)}&\lesssim\|v\|_{W^{1,r}(\omega,\Omega)} &&\forall v\in W^{1,r}(\omega,\Omega),\\
\|v-\Pi_{\mathscr{T}}v\|_{W^{1,r}(\omega,\Omega)}&\lesssim h_{\mathscr{T}}\|v\|_{W^{2,r}(\omega,\Omega)} &&\forall v\in W^{2,r}(\omega,\Omega).
\end{aligned}$$

Finally, to approximate the PDE-constrained optimization problems described in Section 1, we define the space of piecewise constants by

$$\mathbb{U}(\mathscr{T})=\left\{v_{\mathscr{T}}\in L^\infty(\Omega): v_{\mathscr{T}|T}\in\mathbb{P}_0(T)\right\}.\tag{2.4}$$

### 2.3 *Optimality conditions*

To unify the analysis and discretization of the PDE-constrained optimization problems introduced and motivated in Section 1 and thoroughly studied in subsequent sections, we introduce a general framework following the guidelines presented in Lions (1971), Ito & Kunisch (2008), Gamallo & Hernández (2009), Hinze *et al.* (2009), Tröltzsch (2010) and los Reyes (2015). Let $\mathbb{U}$ and $\mathbb{H}$ be Hilbert spaces denoting the so-called control and observation spaces, respectively. We introduce the state trial and test spaces $\mathbb{Y}_1$ and $\mathbb{X}_1$, and the corresponding adjoint test and trial spaces $\mathbb{Y}_2$ and $\mathbb{X}_2$, which we assume to be Hilbert. In addition, we introduce the following items:

(a) a bilinear form $a : (\mathbb{Y}_1 + \mathbb{Y}_2) \times (\mathbb{X}_1 + \mathbb{X}_2) \to \mathbb{R}$ which, when restricted to either $\mathbb{Y}_1 \times \mathbb{X}_1$ or $\mathbb{Y}_2 \times \mathbb{X}_2$, satisfies the conditions of the Banach-Nečas-Babuška (BNB) theorem (see Ern & Guermond, 2004, Theorem 2.6);
(b) a bilinear form $b : \mathbb{U} \times (\mathbb{X}_1 + \mathbb{X}_2) \to \mathbb{R}$ which, when restricted to either $\mathbb{U} \times \mathbb{X}_1$ or $\mathbb{U} \times \mathbb{X}_2$, is bounded (the bilinear forms $a$ and $b$ will be used to describe the state and adjoint equations);
(c) an observation map $C : \mathrm{Dom}(C) \subset \mathbb{Y}_1 + \mathbb{Y}_2 \to \mathbb{H}$, which we assume linear; in addition, we assume that $\mathbb{Y}_2 \subset \mathrm{Dom}(C)$ and that the restriction $C_{|\mathbb{Y}_2} : \mathbb{Y}_2 \to \mathbb{H}$ is continuous;
(d) a desired state $y_d \in \mathbb{H}$;
(e) a regularization parameter $\lambda > 0$ and a cost functional

$$\mathrm{Dom}(C) \times \mathbb{U} \ni (y, u) \mapsto J(y, u) = \frac{1}{2}\|Cy - y_d\|_{\mathbb{H}}^2 + \frac{\lambda}{2}\|u\|_{\mathbb{U}}^2. \tag{2.5}$$

All our problems of interest can be cast as follows. Find $\min J(y, u)$ subject to

$$y \in \mathbb{Y}_1 : \quad a(y, v) = b(u, v) \quad \forall\, v \in \mathbb{X}_1, \tag{2.6}$$

and the constraints

$$u \in \mathbb{U}_{\mathrm{ad}}, \tag{2.7}$$

where $\mathbb{U}_{\mathrm{ad}} \subset \mathbb{U}$ is nonempty, bounded, closed and convex. We introduce the control to state map $S : \mathbb{U} \to \mathbb{Y}_1$ which to a given control, $u \in \mathbb{U}$, associates a unique state, $y(u) = Su \in \mathbb{Y}_1$, that solves the state equation (2.6). As a consequence of (a) and (b), the map $S$ is a bounded and linear operator. If, for every control $u \in \mathbb{U}$, we have $Su \in \mathrm{Dom}(C)$, we can eliminate the state variable $y$ from (2.5) and introduce the reduced cost functional

$$\mathbb{U} \ni u \mapsto j(u) = \frac{1}{2}\|CSu - y_d\|_{\mathbb{H}}^2 + \frac{\lambda}{2}\|u\|_{\mathbb{U}}^2. \tag{2.8}$$

Then, our problem can be cast as follows: find $\min j(u)$ over $\mathbb{U}_{\mathrm{ad}}$. As described in (e) we have $\lambda > 0$ so that $j$ is strictly convex. In addition, $\mathbb{U}_{\mathrm{ad}}$ is weakly sequentially compact in $\mathbb{U}$. Consequently, standard arguments yield existence and uniqueness of a minimizer (Tröltzsch, 2010, Theorem 2.14). In addition, the optimal control $\bar{\mathsf{u}} \in \mathbb{U}_{\mathrm{ad}}$ can be characterized by the variational inequality

$$j'(\bar{\mathsf{u}})(u - \bar{\mathsf{u}}) \geq 0 \quad \forall\, u \in \mathbb{U}_{\mathrm{ad}},$$

where $j'(w)$ denotes the Gâteaux derivative of $j$ at $w$ (Tröltzsch, 2010, Lemma 2.21). Under certain assumptions (see Theorem 2.2 below), this variational inequality can be equivalently written as

$$b(u - \bar{\mathsf{u}}, \bar{\mathsf{p}}) + \lambda(\bar{\mathsf{u}}, u - \bar{\mathsf{u}})_{\mathbb{U}} \geq 0 \quad \forall\, u \in \mathbb{U}_{\mathrm{ad}}, \tag{2.9}$$

where $\bar{\mathsf{p}} = \bar{\mathsf{p}}(\bar{\mathsf{u}})$ denotes the optimal adjoint state and solves

$$\bar{\mathsf{p}} \in \mathbb{X}_2 : \quad a(v, \bar{\mathsf{p}}) = (C\bar{\mathsf{y}} - y_d, Cv)_{\mathbb{H}} \quad \forall\, v \in \mathbb{Y}_2. \tag{2.10}$$

The optimal state $\bar{\mathsf{y}} = \bar{\mathsf{y}}(\bar{\mathsf{u}}) \in \mathbb{Y}_1$ is the solution to (2.6) with $u = \bar{\mathsf{u}}$.

The justification of (2.9)–(2.10) is the content of the next result.

THEOREM 2.2 (Optimality conditions.) Assume that, for every $u \in \mathbb{U}$, we have $Su \in \mathrm{Dom}(C)$. In addition, assume that one of the following two conditions holds:

(i) For every $u \in \mathbb{U}$ we have $Su \in \mathbb{Y}_2$ and there exists $\mathbb{D} \subset \mathbb{X}_1 \cap \mathbb{X}_2$ that is dense in $\mathbb{X}_2$.
(ii) There exists $\mathbb{D} \subset \mathbb{Y}_1 \cap \mathbb{Y}_2$ that is dense in $\mathbb{Y}_1$ and the solution $\bar{\mathsf{p}}$ to (2.10) belongs to $\mathbb{X}_1$. Finally, if $\{y_n\}_{n=1}^{\infty} \subset \mathbb{D}$ is such that, as $n \to \infty$, we have $y_n \to y$ in $\mathbb{Y}_1$, then $Cy_n \to Cy$ in $\mathbb{H}$.

In this setting, the pair $(\bar{\mathsf{y}}, \bar{\mathsf{u}}) \in \mathbb{Y}_1 \times \mathbb{U}$ is optimal if and only if $\bar{\mathsf{y}} = S\bar{\mathsf{u}}$ and $\bar{\mathsf{u}}$ satisfies (2.9), where $\bar{\mathsf{p}} \in \mathbb{X}_2$ is defined by (2.10).

*Proof.* Owing to the particular form of the reduced functional, given in (2.8), the necessary and sufficient condition for optimality reads

$$0 \leq (CS\bar{\mathsf{u}} - y_d, CS(u - \bar{\mathsf{u}}))_{\mathbb{H}} + \lambda(\bar{\mathsf{u}}, u - \bar{\mathsf{u}})_{\mathbb{U}} \quad \forall\, u \in \mathbb{U}_{\mathrm{ad}}.$$

Recall that $\bar{\mathsf{y}} = S\bar{\mathsf{u}}$. To simplify the discussion, set $y = Su$. We now proceed depending on the assumptions:

(i) In this setting, we immediately see, in view of (a) and (c), that (2.10) is well posed and that $v = y - \bar{\mathsf{y}} \in \mathbb{Y}_2$, i.e., $v$ is a valid test function in (2.10). With this particular value of $v$ we get

$$a(y - \bar{\mathsf{y}}, \bar{\mathsf{p}}) = (C\bar{\mathsf{y}} - y_d, C(y - \bar{\mathsf{y}}))_{\mathbb{H}}.$$

Notice that the right-hand side of this expression is the first term on the right-hand side of the variational inequality. By definition of $S$ we have, for every $v_y \in \mathbb{X}_1$,

$$a(y - \bar{\mathsf{y}}, v_y) = b(u - \bar{\mathsf{u}}, v_y). \tag{2.11}$$

In this last identity, we would like to set $v_y = \bar{\mathsf{p}}$ so that we obtain

$$b(u - \bar{\mathsf{u}}, \bar{\mathsf{p}}) = a(y - \bar{\mathsf{y}}, \bar{\mathsf{p}}) = (C\bar{\mathsf{y}} - y_d, C(y - \bar{\mathsf{y}}))_{\mathbb{H}},$$

and this immediately yields (2.9). However $\bar{\mathsf{p}} \notin \mathbb{X}_1$ so we must justify this by a different argument. Let $\{p_n\}_{n=1}^{\infty} \subset \mathbb{D}$ be such that $p_n \to \bar{\mathsf{p}}$ in $\mathbb{X}_2$. Setting $v_y = p_n$ in problem (2.11), which is a valid test function, now yields

$$a(y - \bar{\mathsf{y}}, p_n) = b(u - \bar{\mathsf{u}}, p_n) \to b(u - \bar{\mathsf{u}}, \bar{\mathsf{p}}), \quad n \to \infty,$$

since, by assumption, the form $b$ is continuous on $\mathbb{U}\times\mathbb{X}_2$. On the other hand, the form $a$ is continuous on $\mathbb{Y}_2\times\mathbb{X}_2$ and, since $y-\bar{\mathsf{y}}\in\mathbb{Y}_2$ and $\bar{\mathsf{p}}\in\mathbb{X}_2$, we obtain

$$a(y-\bar{\mathsf{y}},p_n)\to a(y-\bar{\mathsf{y}},\bar{\mathsf{p}}),\quad n\to\infty,$$

which allows us to conclude.

(ii) Under these assumptions we once again obtain that (2.10) is well posed. In addition, since $\bar{\mathsf{p}}\in\mathbb{X}_1$, we can set $v_y=\bar{\mathsf{p}}$ in problem (2.11) to obtain

$$a(y-\bar{\mathsf{y}},\bar{\mathsf{p}})=b(u-\bar{\mathsf{u}},\bar{\mathsf{p}}).$$

The issue at hand now is that setting $v=y-\bar{\mathsf{y}}$ in (2.10) would allow us to conclude. However, $y-\bar{\mathsf{y}}\notin\mathbb{Y}_2$ and so we argue as follows. Let $\{y_n\}_{n=1}^{\infty}\subset\mathbb{D}$ be such that, as $n\to\infty$, it converges to $y-\bar{\mathsf{y}}$ in $\mathbb{Y}_1$. The assumptions then imply that $Cy_n\to C(y-\bar{\mathsf{y}})$ in $\mathbb{H}$. The continuity of $a$ in $\mathbb{Y}_1\times\mathbb{X}_1$ gives

$$a(y_n,\bar{\mathsf{p}})\to a(y-\bar{\mathsf{y}},\bar{\mathsf{p}})=b(u-\bar{\mathsf{u}},\bar{\mathsf{p}}).$$

On the other hand, setting $v=y_n$ in (2.10) yields

$$a(y_n,\bar{\mathsf{p}})=(C\bar{\mathsf{y}}-y_d,Cy_n)_{\mathbb{H}}\to(C\bar{\mathsf{y}}-y_d,C(y-\bar{\mathsf{y}}))_{\mathbb{H}},$$

which allows us to conclude. □

### 2.4 *Discretization of PDE-constrained optimization problems*

Let us now, in the abstract setting of Section 2.3, study the discretization of problem (2.5)–(2.7). Since our ultimate objective is to approximate the problems described in Section 1 with finite element methods, we will study the discretization of (2.5)–(2.7) with Galerkin-like techniques.

Let $h>0$ be a parameter and assume that, for every $h>0$, we have at hand finite-dimensional spaces $\mathbb{U}^h\subset\mathbb{U}$, $\mathbb{X}_1^h\subset\mathbb{X}_1$, $\mathbb{X}_2^h\subset\mathbb{X}_2$, $\mathbb{Y}_1^h\subset\mathbb{Y}_1$ and $\mathbb{Y}_2^h\subset\mathbb{Y}_2$. We define $\mathbb{U}_{\mathrm{ad}}^h=\mathbb{U}^h\cap\mathbb{U}_{\mathrm{ad}}$, which we assume nonempty. About the pairs $(\mathbb{X}_i^h,\mathbb{Y}_i^h)$, for $i=1,2$, we assume that they are such that $a$ satisfies a BNB condition uniformly in $h$ (see Ern & Guermond, 2004, Section 2.2.3). In this setting, the discrete counterpart of (2.5)–(2.7) reads as follows: find

$$\min J(y_h,u_h) \tag{2.12}$$

subject to the discrete state equation

$$y_h\in\mathbb{Y}_1^h:\quad a(y_h,v_h)=b(u_h,v_h)\quad\forall\, v_h\in\mathbb{X}_1^h, \tag{2.13}$$

and the discrete constraints

$$u_h\in\mathbb{U}_{\mathrm{ad}}^h. \tag{2.14}$$

As in the continuous case, we introduce the discrete control to state operator $S_h$, which to a discrete control, $u_h \in \mathbb{U}_h$, associates a unique discrete state, $y_h = y_h(u_h) = S_h u_h$, which solves (2.13). Here $S_h$ is a bounded and linear operator.

The pair $(\bar{\mathsf{y}}_h, \bar{\mathsf{u}}_h) \in \mathbb{Y}_1^h \times \mathbb{U}_{\mathrm{ad}}^h$ is optimal for (2.12)–(2.14) if $\bar{\mathsf{y}}_h = \bar{\mathsf{y}}_h(\bar{\mathsf{u}}_h)$ solves (2.13) and the discrete control $\bar{\mathsf{u}}_h$ satisfies the variational inequality

$$j_h'(\bar{\mathsf{u}}_h)(u_h - \bar{\mathsf{u}}_h) \geq 0 \quad \forall\, u_h \in \mathbb{U}_{\mathrm{ad}}^h,$$

or, under similar assumptions to those of Theorem 2.2, equivalently,

$$b(u_h - \bar{\mathsf{u}}_h, \bar{\mathsf{p}}_h) + \lambda(\bar{\mathsf{u}}_h, u_h - \bar{\mathsf{u}}_h)_{\mathbb{U}} \geq 0 \quad \forall\, u_h \in \mathbb{U}_{\mathrm{ad}}^h, \tag{2.15}$$

where the discrete adjoint variable $\bar{\mathsf{p}}_h = \bar{\mathsf{p}}_h(\bar{\mathsf{u}}_h)$ solves

$$\bar{\mathsf{p}}_h \in \mathbb{X}_2^h: \quad a(v_h, \bar{\mathsf{p}}_h) = (C\bar{\mathsf{y}}_h - y_d, Cv_h)_{\mathbb{H}} \quad \forall\, v_h \in \mathbb{Y}_2^h. \tag{2.16}$$

To develop an error analysis for the discrete problem described above, we introduce $\Pi_{\mathbb{U}}$, the $\mathbb{U}$-orthogonal projection onto $\mathbb{U}^h$. We assume that $\Pi_{\mathbb{U}}\mathbb{U}_{\mathrm{ad}} \subset \mathbb{U}_{\mathrm{ad}}^h$. In addition, we introduce two auxiliary states that will play an important role in the discussion that follows. We define

$$\hat{\mathsf{y}}_h \in \mathbb{Y}_1^h: \quad a(\hat{\mathsf{y}}_h, v_h) = b(\bar{\mathsf{u}}, v_h) \quad \forall\, v_h \in \mathbb{X}_1^h, \tag{2.17}$$

i.e., $\hat{\mathsf{y}}_h$ is defined as the solution to (2.13) with $u_h$ replaced by $\bar{\mathsf{u}}$. We also define

$$\hat{\mathsf{p}}_h \in \mathbb{X}_2^h: \quad a(v_h, \hat{\mathsf{p}}_h) = (C\hat{\mathsf{y}}_h - y_d, Cv_h)_{\mathbb{H}} \quad \forall\, v_h \in \mathbb{Y}_2^h, \tag{2.18}$$

that is, $\hat{\mathsf{p}}_h$ is the solution to (2.16) with $\bar{\mathsf{y}}_h$ replaced by $\hat{\mathsf{y}}_h$.

The main error estimate with this level of abstraction reads as follows.

Lemma 2.3 (Abstract error estimate.) Let $(\bar{\mathsf{y}}, \bar{\mathsf{u}}) \in \mathbb{Y}_1 \times \mathbb{U}_{\mathrm{ad}}$ and $(\bar{\mathsf{y}}_h, \bar{\mathsf{u}}_h) \in \mathbb{Y}_1^h \times \mathbb{U}_{\mathrm{ad}}^h$ be the continuous and discrete optimal pairs that solve (2.5)–(2.7) and (2.12)–(2.14), respectively. If

$$\bar{\mathsf{p}}_h - \hat{\mathsf{p}}_h \in \mathbb{X}_1^h \cap \mathbb{X}_2^h, \qquad \bar{\mathsf{y}}_h - \hat{\mathsf{y}}_h \in \mathbb{Y}_1^h \cap \mathbb{Y}_2^h, \tag{2.19}$$

then, we have the estimate

$$\begin{aligned} \|\bar{\mathsf{u}} - \bar{\mathsf{u}}_h\|_{\mathbb{U}}^2 \lesssim \|\bar{\mathsf{p}} - \hat{\mathsf{p}}_h\|_{\mathbb{X}_2}^2 + j'(\bar{\mathsf{u}})(\Pi_{\mathbb{U}}\bar{\mathsf{u}} - \bar{\mathsf{u}}) + \|\Pi_{\mathbb{U}}\bar{\mathsf{u}} - \bar{\mathsf{u}}\|_{\mathbb{U}}^2 \\ + \left( \sup_{v_p \in \mathbb{Y}_2^h} \frac{(C(\bar{\mathsf{y}}_h - \hat{\mathsf{y}}_h), Cv_p)_{\mathbb{H}}}{\|v_p\|_{\mathbb{Y}_2}} \right)^2, \end{aligned} \tag{2.20}$$

where the hidden constant depends on $\lambda^{-1}$ but does not depend on $h$.

*Proof.* Since by definition $\mathbb{U}_{\mathrm{ad}}^h \subset \mathbb{U}_{\mathrm{ad}}$ and by assumption $\Pi_{\mathbb{U}}\mathbb{U}_{\mathrm{ad}} \subset \mathbb{U}_{\mathrm{ad}}^h$, we set $u = \bar{\mathsf{u}}_h$ and $u_h = \Pi_{\mathbb{U}}\bar{\mathsf{u}}$ in (2.9) and (2.15), respectively. Adding the ensuing inequalities we obtain

$$\lambda\|\bar{\mathsf{u}} - \bar{\mathsf{u}}_h\|_{\mathbb{U}}^2 \leq b(\bar{\mathsf{u}}_h - \bar{\mathsf{u}}, \bar{\mathsf{p}} - \bar{\mathsf{p}}_h) + b(\Pi_{\mathbb{U}}\bar{\mathsf{u}} - \bar{\mathsf{u}}, \bar{\mathsf{p}}_h) + \lambda(\bar{\mathsf{u}}_h, \Pi_{\mathbb{U}}\bar{\mathsf{u}} - \bar{\mathsf{u}})_{\mathbb{U}}. \tag{2.21}$$

Define $\mathrm{I} = b(\bar{\mathsf{u}}_h - \bar{\mathsf{u}}, \bar{\mathsf{p}} - \bar{\mathsf{p}}_h)$. In order to estimate this term, we add and subtract $\hat{\mathsf{p}}_h$ to obtain

$$\mathrm{I} = b(\bar{\mathsf{u}}_h - \bar{\mathsf{u}}, \bar{\mathsf{p}} - \hat{\mathsf{p}}_h) + b(\bar{\mathsf{u}}_h - \bar{\mathsf{u}}, \hat{\mathsf{p}}_h - \bar{\mathsf{p}}_h). \tag{2.22}$$

Since $\hat{\mathsf{p}}_h$ is the unique solution to (2.18), we have

$$a(v_p, \bar{\mathsf{p}}_h - \hat{\mathsf{p}}_h) = (C(\bar{\mathsf{y}}_h - \hat{\mathsf{y}}_h), Cv_p)_{\mathbb{H}} \quad \forall\, v_p \in \mathbb{Y}_2^h. \tag{2.23}$$

Similarly, since $\hat{\mathsf{y}}_h$ solves (2.17), we derive

$$a(\bar{\mathsf{y}}_h - \hat{\mathsf{y}}_h, v_y) = b(\bar{\mathsf{u}}_h - \bar{\mathsf{u}}, v_y) \quad \forall\, v_y \in \mathbb{X}_1^h.$$

Set $v_p = \bar{\mathsf{y}}_h - \hat{\mathsf{y}}_h$ and $v_y = \bar{\mathsf{p}}_h - \hat{\mathsf{p}}_h$. By assumption (2.19), which reads $\bar{\mathsf{p}}_h - \hat{\mathsf{p}}_h \in \mathbb{X}_1^h \cap \mathbb{X}_2^h$ and $\bar{\mathsf{y}}_h - \hat{\mathsf{y}}_h \in \mathbb{Y}_1^h \cap \mathbb{Y}_2^h$, $v_p$ and $v_y$ are admissible test functions. Thus,

$$b(\bar{\mathsf{u}}_h - \bar{\mathsf{u}}, \hat{\mathsf{p}}_h - \bar{\mathsf{p}}_h) = (C(\bar{\mathsf{y}}_h - \hat{\mathsf{y}}_h), C(\hat{\mathsf{y}}_h - \bar{\mathsf{y}}_h))_{\mathbb{H}} \leq 0.$$

This and the continuity of the bilinear form $b$ allow us to bound (2.22) as follows:

$$\mathrm{I} \leq b(\bar{\mathsf{u}}_h - \bar{\mathsf{u}}, \bar{\mathsf{p}} - \hat{\mathsf{p}}_h) \leq \frac{\lambda}{4}\|\bar{\mathsf{u}} - \bar{\mathsf{u}}_h\|_{\mathbb{U}}^2 + \frac{\|b\|^2}{\lambda}\|\bar{\mathsf{p}} - \hat{\mathsf{p}}_h\|_{\mathbb{X}_2}^2,$$

where $\|b\|$ denotes the norm of the bilinear form $b$.

Let us now analyse the remaining terms in (2.21), which we denote by II. To do this, we rewrite II as follows:

$$\begin{aligned}\mathrm{II} = {} & b(\Pi_{\mathbb{U}}\bar{\mathsf{u}} - \bar{\mathsf{u}}, \bar{\mathsf{p}}) + \lambda(\bar{\mathsf{u}}, \Pi_{\mathbb{U}}\bar{\mathsf{u}} - \bar{\mathsf{u}})_{\mathbb{U}} + \lambda(\bar{\mathsf{u}}_h - \bar{\mathsf{u}}, \Pi_{\mathbb{U}}\bar{\mathsf{u}} - \bar{\mathsf{u}})_{\mathbb{U}} \\ & + b(\Pi_{\mathbb{U}}\bar{\mathsf{u}} - \bar{\mathsf{u}}, \hat{\mathsf{p}}_h - \bar{\mathsf{p}}) + b(\Pi_{\mathbb{U}}\bar{\mathsf{u}} - \bar{\mathsf{u}}, \bar{\mathsf{p}}_h - \hat{\mathsf{p}}_h).\end{aligned}$$

Now, notice that

$$b(\Pi_{\mathbb{U}}\bar{\mathsf{u}} - \bar{\mathsf{u}}, \bar{\mathsf{p}}) + \lambda(\bar{\mathsf{u}}, \Pi_{\mathbb{U}}\bar{\mathsf{u}} - \bar{\mathsf{u}})_{\mathbb{U}} = j'(\bar{\mathsf{u}})(\Pi_{\mathbb{U}}\bar{\mathsf{u}} - \bar{\mathsf{u}})$$

and

$$\lambda(\bar{\mathsf{u}}_h - \bar{\mathsf{u}}, \Pi_{\mathbb{U}}\bar{\mathsf{u}} - \bar{\mathsf{u}})_{\mathbb{U}} \leq \frac{\lambda}{4}\|\bar{\mathsf{u}} - \bar{\mathsf{u}}_h\|_{\mathbb{U}}^2 + \frac{1}{\lambda}\|\bar{\mathsf{u}} - \Pi_{\mathbb{U}}\bar{\mathsf{u}}\|_{\mathbb{U}}^2.$$

Next, since the bilinear form $b$ is continuous, we arrive at

$$b(\Pi_{\mathbb{U}}\bar{\mathsf{u}} - \bar{\mathsf{u}}, \hat{\mathsf{p}}_h - \bar{\mathsf{p}}) \leq \frac{\|b\|}{2}\|\Pi_{\mathbb{U}}\bar{\mathsf{u}} - \bar{\mathsf{u}}\|_{\mathbb{U}}^2 + \frac{\|b\|}{2}\|\bar{\mathsf{p}} - \hat{\mathsf{p}}_h\|_{\mathbb{X}_2}^2.$$

The remaining term, which we will denote by III, is treated by using, again, that the bilinear form $b$ is continuous. This implies that

$$\mathrm{III} := b(\Pi_{\mathbb{U}}\bar{\mathsf{u}} - \bar{\mathsf{u}}, \bar{\mathsf{p}}_h - \hat{\mathsf{p}}_h) \leq \frac{\|b\|}{2}\|\Pi_{\mathbb{U}}\bar{\mathsf{u}} - \bar{\mathsf{u}}\|_{\mathbb{U}}^2 + \frac{\|b\|}{2}\|\bar{\mathsf{p}}_h - \hat{\mathsf{p}}_h\|_{\mathbb{X}_2}^2.$$

From (2.23) and the fact that the discrete spaces satisfy a discrete BNB condition uniformly in $h$ we conclude

$$\|\bar{\mathsf{p}}_h - \hat{\mathsf{p}}_h\|_{\mathbb{X}_2} \lesssim \sup_{v_p \in \mathbb{Y}_2^h} \frac{(C(\bar{\mathsf{y}}_h - \hat{\mathsf{y}}_h), Cv_p)_{\mathbb{H}}}{\|v_p\|_{\mathbb{Y}_2^h}}.$$

Collecting these derived estimates we bound the term II.

By placing the estimates that we have obtained for I and II in the inequality (2.21), we arrive at the claimed result. □

The use of this simple result will be illustrated in the following sections.

REMARK 2.4 (Discrete spaces.) In all the examples we will consider below we will have $\mathbb{X}_1^h = \mathbb{X}_2^h = \mathbb{Y}_1^h = \mathbb{Y}_2^h = \mathbb{V}(\mathscr{T})$ algebraically but normed differently, $\mathbb{V}(\mathscr{T})$ being the finite element space defined in (2.3). Consequently, the assumptions of Theorem 2.2 and (2.19) are trivial.

## 3. Optimization with nonuniformly elliptic equations

In this section, we study the problem (1.1)–(1.3) under the abstract framework developed in Section 2.3. Let $\Omega \subset \mathbb{R}^n$ be a convex polytope ($n \geq 1$) and $\omega \in A_2(\mathbb{R}^n)$, where the $A_2$-Muckenhoupt class is given by Definition 2.1. In addition, we assume that $\mathcal{A} : \Omega \to \mathbb{M}^n$ is symmetric and satisfies the nonuniform ellipticity condition (1.4).

### 3.1 *Analysis*

Owing to the fact that the diffusion matrix $\mathcal{A}$ satisfies (1.4) with $\omega \in A_2(\mathbb{R}^n)$, as shown in Fabes *et al.* (1982), the state equation (1.2) is well posed in $H_0^1(\omega, \Omega)$, whenever $u \in L^2(\omega^{-1}, \Omega)$. For this reason, we set

- $\mathbb{H} = L^2(\omega, \Omega)$ and $C = \mathrm{id}$;
- $\mathbb{U} = L^2(\omega^{-1}, \Omega)$;
- $\mathbb{X}_1 = \mathbb{X}_2 = \mathbb{Y}_1 = \mathbb{Y}_2 = H_0^1(\omega, \Omega)$, and

$$a(v_1, v_2) = \int_\Omega \nabla v_2(x)^{\mathrm{T}} \mathcal{A}(x) \nabla v_1(x)\, \mathrm{d}x,$$

  which, as a consequence of (1.4) with $\omega \in A_2(\mathbb{R}^n)$ and the Poincaré inequality (2.2), is bounded, symmetric and coercive in $H_0^1(\omega, \Omega)$;
- $b(\cdot,\cdot) = (\cdot,\cdot)_{L^2(\Omega)}$; notice that, if $v_1 \in L^2(\omega^{-1}, \Omega)$ and $v_2 \in H_0^1(\omega, \Omega)$ then

$$b(v_1, v_2) = (v_1, v_2)_{L^2(\Omega)} \leq \|v_1\|_{L^2(\omega^{-1},\Omega)} \|v_2\|_{L^2(\omega,\Omega)} \lesssim \|v_1\|_{L^2(\omega^{-1},\Omega)} \|\nabla v_2\|_{L^2(\omega,\Omega)},$$

  where we have used the Poincaré inequality (2.2);
- the cost functional as in (1.1).

For $\mathsf{a}, \mathsf{b} \in \mathbb{R}$, $\mathsf{a} < \mathsf{b}$, we define the set of admissible controls by

$$\mathbb{U}_{\mathcal{A}} = \left\{ u \in L^2(\omega^{-1}, \Omega) : \mathsf{a} \le u \le \mathsf{b} \text{ a.e. } x \in \Omega \right\}, \tag{3.1}$$

which is closed, bounded and convex in $L^2(\omega^{-1}, \Omega)$. In addition, since $\lambda > 0$ the functional (1.1) is strictly convex. Consequently, the optimization problem with nonuniformly elliptic state equation (1.1)–(1.3) has a unique optimal pair $(\bar{\mathsf{y}}, \bar{\mathsf{u}}) \in H_0^1(\omega, \Omega) \times L^2(\omega^{-1}, \Omega)$ (Tröltzsch, 2010, Theorem 2.14). Notice that, in this setting, the conditions of Theorem 2.2(i) are trivially satisfied. In fact, set $\mathbb{D} = C_0^\infty(\Omega)$ and notice that, for $u \in L^2(\omega^{-1}, \Omega)$, we have $Su \in \mathbb{Y}_1 = \mathbb{Y}_2 \subset \mathrm{Dom}(C) = L^2(\omega, \Omega)$. Consequently, the first-order necessary and sufficient optimality condition (2.9) reads

$$(\bar{\mathsf{p}}, u - \bar{\mathsf{u}})_{L^2(\Omega)} + \lambda (\bar{\mathsf{u}}, u - \bar{\mathsf{u}})_{L^2(\omega^{-1}, \Omega)} \ge 0 \quad \forall\, u \in \mathbb{U}_{\mathcal{A}}, \tag{3.2}$$

where the optimal state $\bar{\mathsf{y}} = \bar{\mathsf{y}}(\bar{\mathsf{u}}) \in H_0^1(\omega, \Omega)$ solves

$$a(\bar{\mathsf{y}}, v) = (\bar{\mathsf{u}}, v)_{L^2(\Omega)} \quad \forall\, v \in H_0^1(\omega, \Omega) \tag{3.3}$$

and the optimal adjoint state $\bar{\mathsf{p}} = \bar{\mathsf{p}}(\bar{\mathsf{u}}) \in H_0^1(\omega, \Omega)$ solves

$$a(v, \bar{\mathsf{p}}) = (\bar{\mathsf{y}} - y_d, v)_{L^2(\omega, \Omega)} \quad \forall\, v \in H_0^1(\omega, \Omega). \tag{3.4}$$

The results of Fabes *et al.* (1982), again, yield that the adjoint problem is well posed.

### 3.2 *Discretization*

Let us now propose a discretization for problem (1.1)–(1.3) and derive *a priori* error estimates based on the results of Section 2.4. Given a family $\mathbb{T} = \{\mathscr{T}\}$ of quasi-uniform triangulations of $\Omega$ we set

- $\mathbb{U}^h = \mathbb{U}(\mathscr{T})$, where the discrete space $\mathbb{U}(\mathscr{T})$ is defined in (2.4);
- $\mathbb{U}^h_{\mathrm{ad}} = \mathbb{U}^h \cap \mathbb{U}_{\mathcal{A}}$, where the set of admissible controls $\mathbb{U}_{\mathcal{A}}$ is defined in (3.1);
- $\Pi_{\mathbb{U}}$ is the $L^2(\omega^{-1}, \Omega)$-orthogonal projection onto $\mathbb{U}(\mathscr{T})$, which we denote by $\Pi_{\omega^{-1}}$ and is defined by

$$(\Pi_{\omega^{-1}} v)_{|T} = \frac{1}{\omega^{-1}(T)} \int_T \omega^{-1}(x) v(x)\, \mathrm{d}x \quad \forall\, T \in \mathscr{T}, \tag{3.5}$$

  where $\omega^{-1}(T)$ is defined as in (2.1); the definition of $\mathbb{U}_{\mathcal{A}}$ yields that $\Pi_{\omega^{-1}} \mathbb{U}_{\mathcal{A}} \subset \mathbb{U}^h_{\mathrm{ad}}$;
- $\mathbb{X}^h_1 = \mathbb{X}^h_2 = \mathbb{Y}^h_1 = \mathbb{Y}^h_2 = \mathbb{V}(\mathscr{T})$, where the discrete space $\mathbb{V}(\mathscr{T})$ is defined in (2.3).

Notice that, since $\mathbb{X}^h_1 = \mathbb{X}^h_2 = \mathbb{Y}^h_1 = \mathbb{Y}^h_2$, the assumptions of Theorem 2.2 and (2.19) are trivially satisfied; see Remark 2.4.

We obtain the following *a priori* error estimate.

COROLLARY 3.1 (*A priori* error estimate) Let $\bar{u}$ and $\bar{u}_h$ be the continuous and discrete optimal controls, respectively. If $\bar{y}, \bar{p} \in H^2(\omega, \Omega)$ then

$$\|\bar{u} - \bar{u}_h\|_{L^2(\omega^{-1},\Omega)} \lesssim \left\|\bar{u} - \Pi_{\omega^{-1}}\bar{u}\right\|_{L^2(\omega^{-1},\Omega)} + \|\omega\bar{p} - \Pi_{\omega^{-1}}(\omega\bar{p})\|_{L^2(\omega^{-1},\Omega)} + h_{\mathscr{T}}(\|\bar{y}\|_{H^2(\omega,\Omega)} + \|\bar{p}\|_{H^2(\omega,\Omega)}),$$

where the hidden constant is independent of $h_{\mathscr{T}}$.

*Proof.* We invoke Lemma 2.3 and bound each one of the terms in (2.20). First, since $\bar{y}, \bar{p} \in H^2(\omega, \Omega)$, the results of Nochetto *et al.* (2016) imply that

$$\|\bar{p} - \hat{p}_h\|_{H^1(\omega,\Omega)} \lesssim h_{\mathscr{T}} \left(\|\bar{y}\|_{H^2(\omega,\Omega)} + \|\bar{p}\|_{H^2(\omega,\Omega)}\right).$$

Indeed, since $\bar{p}$ solves (3.4) and $\hat{p}_h$ solves (2.18), the term $\bar{p} - \hat{p}_h$ satisfies

$$a(v_h, \bar{p} - \hat{p}_h) = (\bar{y} - \hat{y}_h, v_h)_{L^2(\omega,\Omega)} \quad \forall\, v_h \in \mathbb{V}(\mathscr{T}).$$

Adding and subtracting the terms $\Pi_{\mathscr{T}}\bar{p}$ and $\bar{p}$ appropriately, where $\Pi_{\mathscr{T}}$ denotes the interpolation operator described in Section 2.2, and using the coercivity of $a$ we arrive at

$$\|\bar{p} - \hat{p}_h\|_{H_0^1(\omega,\Omega)} \lesssim \|\bar{p} - \Pi_{\mathscr{T}}\bar{p}\|_{H_0^1(\omega,\Omega)} + \|\bar{y} - \hat{y}_h\|_{H_0^1(\omega,\Omega)}.$$

Using the regularity of $\bar{p}$ and $\bar{y}$ we obtain the claimed bound.

We now handle the second term involving the derivative of the reduced cost $j$. Since it can be equivalently written using (2.9), invoking the definition of $\Pi_{\omega^{-1}}$ given by (3.5), we obtain

$$\begin{aligned} j'(\bar{u})(\Pi_{\omega^{-1}}\bar{u} - \bar{u}) &= (\bar{p}, \Pi_{\omega^{-1}}\bar{u} - \bar{u})_{L^2(\Omega)} + \lambda(\bar{u}, \Pi_{\omega^{-1}}\bar{u} - \bar{u})_{L^2(\omega^{-1},\Omega)} \\ &= (\omega\bar{p} - \Pi_{\omega^{-1}}(\omega\bar{p}), \Pi_{\omega^{-1}}\bar{u} - \bar{u})_{L^2(\omega^{-1},\Omega)} - \lambda\|\Pi_{\omega^{-1}}\bar{u} - \bar{u}\|^2_{L^2(\omega^{-1},\Omega)} \\ &\lesssim \|\omega\bar{p} - \Pi_{\omega^{-1}}(\omega\bar{p})\|^2_{L^2(\omega^{-1},\Omega)} + \|\Pi_{\omega^{-1}}\bar{u} - \bar{u}\|^2_{L^2(\omega^{-1},\Omega)}. \end{aligned}$$

The Poincaré inequality (2.2), in conjunction with the stability of the discrete state equation (2.13), yields

$$\begin{aligned} (\bar{y}_h - \hat{y}_h, v_h)_{L^2(\omega,\Omega)} &\lesssim \|\bar{y}_h - \hat{y}_h\|_{H_0^1(\omega,\Omega)}\|v_h\|_{H_0^1(\omega,\Omega)} \\ &\lesssim \|\bar{u} - \bar{u}_h\|_{L^2(\omega^{-1},\Omega)}\|v_h\|_{H_0^1(\omega,\Omega)} \end{aligned}$$

for all $v_h \in \mathbb{V}(\mathscr{T})$. This yields control of the last term in (2.20).

These bounds yield the result. □

REMARK 3.2 (Regularity of $\bar{y}$ and $\bar{p}$.) The results of Corollary 3.1 rely on the fact that $\bar{y}, \bar{p} \in H^2(\omega, \Omega)$. Reference Cavalheiro (2011) provides sufficient conditions for this to hold.

THEOREM 3.3 (Rate of convergence.) In the setting of Corollary 3.1, if we additionally assume that $\omega\bar{\mathsf{p}} \in H^1(\omega^{-1}, \Omega)$ then we have the optimal error estimate

$$\|\bar{\mathsf{u}} - \bar{\mathsf{u}}_h\|_{L^2(\omega^{-1},\Omega)} \lesssim h_{\mathscr{T}} \left( \|\bar{\mathsf{y}}\|_{H^2(\omega,\Omega)} + \|\bar{\mathsf{p}}\|_{H^2(\omega,\Omega)} + \|\omega\bar{\mathsf{p}}\|_{H^1(\omega^{-1},\Omega)} + \|\bar{\mathsf{u}}\|_{H^1(\omega^{-1},\Omega)} \right),$$

where the hidden constant is independent of $h_{\mathscr{T}}$.

*Proof.* We bound $\left\|\bar{\mathsf{u}} - \Pi_{\omega^{-1}}\bar{\mathsf{u}}\right\|_{L^2(\omega^{-1},\Omega)}$ and $\|\omega\bar{\mathsf{p}} - \Pi_{\omega^{-1}}(\omega\bar{\mathsf{p}})\|_{L^2(\omega^{-1},\Omega)}$. Using that $\omega\bar{\mathsf{p}} \in H^1(\omega^{-1}, \Omega)$ and a Poincaré-type inequality (Nochetto *et al.*, 2016, Theorem 6.2), we obtain

$$\|\omega\bar{\mathsf{p}} - \Pi_{\omega^{-1}}(\omega\bar{\mathsf{p}})\|_{L^2(\omega^{-1},\Omega)} \lesssim h_{\mathscr{T}} \|\omega\bar{\mathsf{p}}\|_{H^1(\omega^{-1},\Omega)}.$$

Now, to estimate the term $\Pi_{\omega^{-1}}\bar{\mathsf{u}} - \bar{\mathsf{u}}$, it is essential to understand the regularity properties of $\bar{\mathsf{u}}$. From Tröltzsch (2010, Section 3.6.3), $\bar{\mathsf{u}}$ solves (3.2) if and only if

$$\bar{\mathsf{u}} = \max\left\{\mathsf{a}, \min\left\{\mathsf{b}, -\frac{1}{\lambda}\omega\bar{\mathsf{p}}\right\}\right\}.$$

The assumption $\omega\bar{\mathsf{p}} \in H^1(\omega^{-1}, \Omega)$ immediately yields $\bar{\mathsf{u}} \in H^1(\omega^{-1}, \Omega)$ (Kinderlehrer & Stampacchia, 1980, Theorem A.1), which allows us to derive the estimate

$$\|\bar{\mathsf{u}} - \Pi_{\omega^{-1}}\bar{\mathsf{u}}\|_{L^2(\omega^{-1},\Omega)} \lesssim h_{\mathscr{T}} \|\bar{\mathsf{u}}\|_{H^1(\omega^{-1},\Omega)}.$$

Collecting the derived results we arrive at the desired estimate. □

## 4. Optimization with point observations

Here, we consider problem (1.5)–(1.7). Let $\Omega \subset \mathbb{R}^n$ be a convex polytope, with $n \in \{2, 3\}$. We recall that $\mathcal{Z} \subset \Omega$ denotes the set of *observable points* with $\#\mathcal{Z} < \infty$.

### 4.1 *Analysis*

To analyse problem (1.5)–(1.7) using the framework of weighted spaces, we must begin by defining a suitable weight. If $\#\mathcal{Z} = 1$, define $d_{\mathcal{Z}} = 1/2$, otherwise, since $\#\mathcal{Z} < \infty$, we set $d_{\mathcal{Z}} = \min\{|z - z'| : z, z' \in \mathcal{Z},\ z \neq z'\} > 0$. For each $z \in \mathcal{Z}$, we then define

$$\mathsf{d}_z(x) = \frac{1}{2d_{\mathcal{Z}}}|x - z|, \qquad \varpi_z(x) = \frac{\mathsf{d}_z(x)^{n-2}}{\log^2 \mathsf{d}_z(x)}$$

and the weight

$$\varpi(x) = \varpi_z(x),\ \#\mathcal{Z} = 1, \quad \varpi(x) = \begin{cases} \varpi_z(x), & \exists z \in \mathcal{Z} : \mathsf{d}_z(x) < \frac{1}{2}, \\ \dfrac{2^{2-n}}{\log^2 2} & \text{otherwise,} \end{cases} \quad \#\mathcal{Z} > 1. \tag{4.1}$$

As Nochetto *et al.* (2016, Lemma 7.5) and Aimar *et al.* (2014) show, with this definition we have $\varpi \in A_2$. With this $A_2$ weight at hand we set

- $\mathbb{H} = \mathbb{R}^{\#\mathcal{Z}}$ and $C = \sum_{z\in\mathcal{Z}} \mathbf{e}_z\delta_z$, where $\{\mathbf{e}_z\}_{z\in\mathcal{Z}}$ is the canonical basis of $\mathbb{H}$;
- $\mathbb{U} = L^2(\Omega)$;
- $\mathbb{X}_1 = \mathbb{Y}_1 = H_0^1(\Omega)$;
- $\mathbb{X}_2 = H_0^1(\varpi, \Omega)$ and $\mathbb{Y}_2 = H_0^1(\varpi^{-1}, \Omega)$ and

$$a(v, w) = \int_\Omega \nabla v(x)^{\mathrm{T}} \cdot \nabla w(x)\,\mathrm{d}x,$$

  which is bounded, symmetric and coercive in $H_0^1(\Omega)$ and satisfies the conditions of the BNB theorem in $H_0^1(\varpi, \Omega) \times H_0^1(\varpi^{-1}, \Omega)$ (Nochetto *et al.*, 2016, Lemma 7.7);
- $b(\cdot,\cdot) = (\cdot,\cdot)_{L^2(\Omega)}$. The results of Nochetto *et al.* (2016, Lemma 7.6) guarantee that, for $n < 4$, the embedding $H_0^1(\varpi, \Omega) \hookrightarrow L^2(\Omega)$ holds; therefore,

$$b(v_1, v_2) \lesssim \|v_1\|_{L^2(\Omega)}\|v_2\|_{H_0^1(\varpi,\Omega)}.$$

For $\mathsf{a}, \mathsf{b} \in \mathbb{R}$, with $\mathsf{a} < \mathsf{b}$ we define the set of admissible controls by

$$\mathbb{U}_{\mathcal{Z}} = \left\{u \in L^2(\Omega) : \mathsf{a} \le u \le \mathsf{b},\ a.e.\ x \in \Omega\right\}. \tag{4.2}$$

With this notation, the pair $(\bar{\mathsf{y}}, \bar{\mathsf{u}}) \in H_0^1(\Omega) \times L^2(\Omega)$ is optimal for problem (1.5)–(1.7) if and only if $\bar{\mathsf{y}}$ solves

$$\bar{\mathsf{y}} \in H_0^1(\Omega) : \quad a(\bar{\mathsf{y}}, w) = (\bar{\mathsf{u}}, w)_{L^2(\Omega)} \quad \forall\, w \in H_0^1(\Omega), \tag{4.3}$$

and the optimal control $\bar{\mathsf{u}}$ satisfies

$$(\bar{\mathsf{p}}, u - \bar{\mathsf{u}})_{L^2(\Omega)} + \lambda(\bar{\mathsf{u}}, u - \bar{\mathsf{u}})_{L^2(\Omega)} \ge 0 \quad \forall\, u \in \mathbb{U}_{\mathcal{Z}}, \tag{4.4}$$

where the adjoint variable $\bar{\mathsf{p}} \in H_0^1(\varpi, \Omega)$ satisfies, for every $w \in H_0^1(\varpi^{-1}, \Omega)$,

$$a(w, \bar{\mathsf{p}}) = \sum_{z\in\mathcal{Z}} (\bar{\mathsf{y}}(z) - y_z)\langle\delta_z, w\rangle_{H_0^1(\varpi^{-1},\Omega)'\times H_0^1(\varpi^{-1},\Omega)}. \tag{4.5}$$

Indeed, it suffices to set, in Theorem 2.2, $\mathbb{D} = C_0^\infty(\Omega)$ and to recall that since $\Omega$ is a convex polytope and $n < 4$, we have $\bar{\mathsf{y}} \in H^2(\Omega) \hookrightarrow C(\bar{\Omega})$, so point evaluations are meaningful, i.e., $y = Su \in \mathrm{Dom}(C)$. Finally, the embedding of Nochetto *et al.* (2016, Lemma 7.6) shows that $\bar{\mathsf{y}} \in H^2(\Omega) \cap H_0^1(\Omega) \hookrightarrow H_0^1(\varpi^{-1}, \Omega) = \mathbb{Y}_2$, that is, item (i) is satisfied. In addition, since $\delta_z \in H_0^1(\varpi^{-1}, \Omega)'$ for $z \in \Omega$, we thus have $H_0^1(\varpi^{-1}, \Omega) = \mathbb{Y}_2 \subset \mathrm{Dom}(C)$ and, in view of Nochetto *et al.* (2016, Lemma 7.7), the adjoint problem is well posed.

### 4.2 *Discretization*

For a family $\mathbb{T} = \{\mathscr{T}\}$ of quasi-uniform meshes of $\Omega$ we set

- $\mathbb{U}^h = \mathbb{U}(\mathscr{T})$, where $\mathbb{U}(\mathscr{T})$ is defined in (2.4) and $\mathbb{U}^h_{\rm ad} = \mathbb{U}(\mathscr{T}) \cap \mathbb{U}_{\mathcal{Z}}$, where $\mathbb{U}_{\mathcal{Z}}$ is defined in (4.2). The operator $\Pi_{\mathbb{U}} = \Pi_{L^2}$ is the standard $L^2(\Omega)$-projection:

$$(\Pi_{L^2} v)_{|T} = \fint_T v(x)\,\mathrm{d}x \quad \forall\, T \in \mathscr{T}.$$

- $\mathbb{X}^h_1 = \mathbb{X}^h_2 = \mathbb{Y}^h_1 = \mathbb{Y}^h_2 = \mathbb{V}(\mathscr{T})$. As before, this makes the assumptions of Theorem 2.2 and (2.19) trivial.

To shorten the exposition, we define

$$\sigma_{\mathscr{T}} = h_{\mathscr{T}}^{2-n/2} |\log h_{\mathscr{T}}|. \tag{4.6}$$

With this notation, the error estimate for the approximation (2.12)–(2.14) to problem (1.5)–(1.7) reads as follows.

COROLLARY 4.1 (*A priori* error estimates.) Let $\bar{\mathsf{u}}$ and $\bar{\mathsf{u}}_h$ be the continuous and discrete optimal controls, respectively. Assume that $h_{\mathscr{T}}$ is sufficiently small. Then, for $n \in \{2, 3\}$, we have the error estimate

$$\|\bar{\mathsf{u}} - \bar{\mathsf{u}}_h\|_{L^2(\Omega)} \lesssim \|\bar{\mathsf{u}} - \Pi_{L^2}\bar{\mathsf{u}}\|_{L^2(\Omega)} + \sigma_{\mathscr{T}} \left( \|\nabla \bar{\mathsf{p}}\|_{L^2(\varpi,\Omega)} + \|\nabla \bar{\mathsf{y}}\|_{L^\infty(\Omega)} \right), \tag{4.7}$$

where $\sigma_{\mathscr{T}}$ is defined in (4.6) and the hidden constant is independent of $\mathscr{T}$.

*Proof.* We follow Lemma 2.3 with slight modifications. The term I in (2.22) is estimated in two steps. In fact, since $(\bar{\mathsf{u}}_h - \bar{\mathsf{u}}, \hat{\mathsf{p}}_h - \bar{\mathsf{p}}_h)_{L^2(\Omega)} \le 0$, we have

$$\mathrm{I} \le (\bar{\mathsf{u}}_h - \bar{\mathsf{u}}, \bar{\mathsf{p}} - \hat{\mathsf{p}}_h)_{L^2(\Omega)} \le \frac{\lambda}{4}\|\bar{\mathsf{u}}_h - \bar{\mathsf{u}}\|^2_{L^2(\Omega)} + \frac{1}{\lambda}\|\bar{\mathsf{p}} - \hat{\mathsf{p}}_h\|^2_{L^2(\Omega)}.$$

We now analyse the second term of the previous expression. Let $\mathsf{q}_h \in \mathbb{V}(\mathscr{T})$ solve

$$a(w_h, \mathsf{q}_h) = \sum_{z \in \mathcal{Z}} (\bar{\mathsf{y}}(z) - y_z) w_h(z) \quad \forall\, w_h \in \mathbb{V}(\mathscr{T}). \tag{4.8}$$

The conclusion of Nochetto *et al.* (2016, Corollary 7.9) immediately yields

$$\|\bar{\mathsf{p}} - \mathsf{q}_h\|_{L^2(\Omega)} \lesssim \sigma_{\mathscr{T}} \|\nabla \bar{\mathsf{p}}\|_{L^2(\varpi,\Omega)},$$

so that it remains to estimate $\mathsf{q}_h - \hat{\mathsf{p}}_h$. We now invoke Nochetto *et al.* (2016, Theorem 6.1) with $p = q = 2$, $\rho = 1$ and $\omega = \varpi$. Under this setting, the compatibility condition Nochetto *et al.* (2016, inequality (6.2)) is satisfied, and then Nochetto *et al.* (2016, Theorem 6.1) yields

$$\|\mathsf{q}_h - \hat{\mathsf{p}}_h\|_{L^2(\Omega)} \lesssim \|\nabla(\mathsf{q}_h - \hat{\mathsf{p}}_h)\|_{L^2(\varpi,\Omega)},$$

where the hidden constant depends on $\Omega$, the quotient between the radii of the balls inscribed and circumscribed in $\Omega$ and the weight $\varpi$ only through the constant $\varpi(\Omega)$; the latter is defined as in (2.1). Since $\mathsf{q}_h$ solves (4.8), the discrete inf–sup conditions of Nochetto *et al.* (2016, Lemma 7.8) and the fact that $\delta_z \in H_0^1(\varpi^{-1}, \Omega)'$ yield

$$\|\nabla(\mathsf{q}_h - \hat{\mathsf{p}}_h)\|_{L^2(\varpi,\Omega)} \lesssim \|\bar{\mathsf{y}} - \hat{\mathsf{y}}_h\|_{L^\infty(\Omega)}.$$

We now recall that $\hat{\mathsf{y}}_h$ is the Galerkin projection of $\bar{\mathsf{y}}$. In addition, since $n \in \{2, 3\}$, $\Omega$ is a convex polytope and $\bar{\mathsf{u}} \in L^\infty(\Omega)$, we have $\bar{\mathsf{y}} \in W^{1,\infty}(\Omega)$ (cf. Maz′ya & Rossmann, 1991; Fromm, 1993; Guzmán *et al.*, 2009). Therefore, standard pointwise estimates for finite elements (Schatz & Wahlbin, 1982, Theorem 5.1) yield

$$\|\bar{\mathsf{y}} - \hat{\mathsf{y}}_h\|_{L^\infty(\Omega)} \lesssim h_{\mathscr{T}} |\log h_{\mathscr{T}}| \|\nabla\bar{\mathsf{y}}\|_{L^\infty(\Omega)}. \tag{4.9}$$

In conclusion,

$$\mathrm{I} \leq \frac{\lambda}{4}\|\bar{\mathsf{u}}_h - \bar{\mathsf{u}}\|^2_{L^2(\Omega)} + c\sigma^2_{\mathscr{T}}\left(\|\nabla\bar{\mathsf{p}}\|^2_{L^2(\varpi,\Omega)} + \|\nabla\bar{\mathsf{y}}\|^2_{L^\infty(\Omega)}\right),$$

for some nonessential constant $c$.

We estimate the term $j'(\bar{\mathsf{u}})(\Pi_{L^2}\bar{\mathsf{u}} - \bar{\mathsf{u}})$ as follows:

$$\begin{aligned} j'(\bar{\mathsf{u}})(\Pi_{L^2}\bar{\mathsf{u}} - \bar{\mathsf{u}}) &= (\bar{\mathsf{p}} + \lambda\bar{\mathsf{u}}, \Pi_{L^2}\bar{\mathsf{u}} - \bar{\mathsf{u}})_{L^2(\Omega)} = (\bar{\mathsf{p}} + \lambda\bar{\mathsf{u}} - \Pi_{L^2}(\bar{\mathsf{p}} + \lambda\bar{\mathsf{u}}), \Pi_{L^2}\bar{\mathsf{u}} - \bar{\mathsf{u}})_{L^2(\Omega)} \\ &\leq \frac{1}{2}\|\Pi_{L^2}\bar{\mathsf{u}} - \bar{\mathsf{u}}\|^2_{L^2(\Omega)} + \frac{1}{2}\|\bar{\mathsf{p}} - \Pi_{L^2}\bar{\mathsf{p}}\|^2_{L^2(\Omega)} \\ &\leq \frac{1}{2}\|\Pi_{L^2}\bar{\mathsf{u}} - \bar{\mathsf{u}}\|^2_{L^2(\Omega)} + c\sigma^2_{\mathscr{T}}\|\nabla\bar{\mathsf{p}}\|^2_{L^2(\varpi,\Omega)}, \end{aligned}$$

for some nonessential constant $c$. We have used the properties of $\Pi_{L^2}$, together with the Sobolev–Poincaré inequality of Nochetto *et al.* (2016, Theorem 6.2); see also Nochetto *et al.* (2016, Corollary 7.9).

We now proceed to estimate the term III in the proof of Lemma 2.3 as follows:

$$\begin{aligned} \mathrm{III} := b(\Pi_{L^2}\bar{\mathsf{u}} - \bar{\mathsf{u}}, \bar{\mathsf{p}}_h - \hat{\mathsf{p}}_h) &= (\Pi_{L^2}\bar{\mathsf{u}} - \bar{\mathsf{u}}, \bar{\mathsf{p}}_h - \hat{\mathsf{p}}_h)_{L^2(\Omega)} \\ &\leq \frac{1}{2}\|\Pi_{L^2}\bar{\mathsf{u}} - \bar{\mathsf{u}}\|^2_{L^2(\Omega)} + \frac{1}{2}\|\bar{\mathsf{p}}_h - \hat{\mathsf{p}}_h - \Pi_{L^2}(\bar{\mathsf{p}}_h - \hat{\mathsf{p}}_h)\|^2_{L^2(\Omega)} \\ &\leq \frac{1}{2}\|\Pi_{L^2}\bar{\mathsf{u}} - \bar{\mathsf{u}}\|^2_{L^2(\Omega)} + c\sigma^2_{\mathscr{T}}\|\nabla(\bar{\mathsf{p}}_h - \hat{\mathsf{p}}_h)\|^2_{L^2(\varpi,\Omega)}, \end{aligned}$$

where we have used the properties of $\Pi_{L^2}$ together with the Sobolev–Poincaré inequality of Nochetto *et al.* (2016, Theorem 6.2); again, $c$ denotes a nonessential constant. Using now the fact that $\delta_z \in H_0^1(\varpi^{-1}, \Omega)'$, and the discrete inf–sup stability of Nochetto *et al.* (2016, Lemma 7.8), we have

$$\|\bar{\mathsf{p}}_h - \hat{\mathsf{p}}_h\|_{H_0^1(\varpi,\Omega)} \lesssim \|\bar{\mathsf{y}}_h - \hat{\mathsf{y}}_h\|_{L^\infty(\Omega)}. \tag{4.10}$$

To estimate the right-hand side of the previous expression, we introduce the state

$$y(\bar{\mathsf{u}}_h) \in H_0^1(\Omega): \quad a(y(\bar{\mathsf{u}}_h), w) = (\bar{\mathsf{u}}_h, w)_{L^2(\Omega)} \quad \forall\, w \in H_0^1(\Omega),$$

and define $\chi := \bar{\mathsf{y}} - y(\bar{\mathsf{u}}_h) \in H_0^1(\Omega)$ and $\chi_h := \hat{\mathsf{y}}_h - \bar{\mathsf{y}}_h \in \mathbb{V}(\mathscr{T})$, where $\bar{\mathsf{y}}_h$ and $\hat{\mathsf{y}}_h$ solve (2.13) and (2.17), respectively. We observe that

$$a(\chi, w) = (\bar{\mathsf{u}} - \bar{\mathsf{u}}_h, w)_{L^2(\Omega)} \quad \forall\, w \in H_0^1(\Omega),$$
$$a(\chi_h, w_h) = (\bar{\mathsf{u}} - \bar{\mathsf{u}}_h, w_h)_{L^2(\Omega)} \quad \forall\, w_h \in \mathbb{V}(\mathscr{T}),$$

i.e., $\chi_h$ is the Galerkin approximation of $\chi$. If we denote by $I_h : C(\bar{\Omega}) \to \mathbb{V}(\mathscr{T})$ the Lagrange interpolation operator (Ern & Guermond, 2004), basic applications of the triangle inequality and a standard inverse estimate (Ern & Guermond, 2004, Lemma 1.138) yield

$$\begin{aligned}\|\chi_h\|_{L^\infty(\Omega)} &\lesssim \|\chi\|_{L^\infty(\Omega)} + \|\chi - I_h\chi\|_{L^\infty(\Omega)} + h^{-n/2}\|I_h\chi - \chi_h\|_{L^2(\Omega)} \\ &\lesssim \|\chi\|_{L^\infty(\Omega)} + \|\chi - I_h\chi\|_{L^\infty(\Omega)} + h^{-n/2}\left(\|\chi - \chi_h\|_{L^2(\Omega)} + \|\chi - I_h\chi\|_{L^2(\Omega)}\right). \end{aligned} \tag{4.11}$$

To control the first term on the right-hand side of (4.11), we invoke the results of Jerison & Kenig (1995, Theorem 0.5) to conclude that there is $r > n$ such that

$$\|\chi\|_{W^{1,r}(\Omega)} \lesssim \|\bar{\mathsf{u}} - \bar{\mathsf{u}}_h\|_{L^2(\Omega)}.$$

This, in view of the continuous embedding $W^{1,r}(\Omega) \hookrightarrow C(\bar{\Omega})$ for $r > n$, yields

$$\|\chi\|_{L^\infty(\Omega)} \lesssim \|\bar{\mathsf{u}} - \bar{\mathsf{u}}_h\|_{L^2(\Omega)}. \tag{4.12}$$

The second and third terms on the right-hand side of (4.11) are bounded in view of standard interpolation and error estimates. We thus have

$$\|\chi_h\|_{L^\infty(\Omega)} \lesssim \|\bar{\mathsf{u}} - \bar{\mathsf{u}}_h\|_{L^2(\Omega)}. \tag{4.13}$$

From (4.10), (4.13) and the fact that $\chi_h = \hat{\mathsf{y}}_h - \bar{\mathsf{y}}_h$, it follows that

$$\|\bar{\mathsf{p}}_h - \hat{\mathsf{p}}_h\|_{H_0^1(\varpi,\Omega)} \lesssim \|\bar{\mathsf{u}} - \bar{\mathsf{u}}_h\|_{L^2(\Omega)}.$$

Therefore, we have derived the following estimate for the term III:

$$\mathrm{III} \le \frac{1}{2}\|\Pi_{L^2}\bar{\mathsf{u}} - \bar{\mathsf{u}}\|^2_{L^2(\Omega)} + c\sigma^2_{\mathscr{T}}\|\bar{\mathsf{u}} - \bar{\mathsf{u}}_h\|^2_{L^2(\Omega)}.$$

Collecting the derived estimates for the terms I, II and III, we arrive at the desired estimate (4.7) by considering $h_{\mathscr{T}}$ sufficiently small. □

PROPOSITION 4.2 (Regularity of $\bar{\mathsf{u}}$.) If $\bar{\mathsf{u}}$ solves (1.5)–(1.7) then $\bar{\mathsf{u}} \in H^1(\varpi, \Omega)$.

*Proof.* From Tröltzsch (2010, Section 3.6.3), $\bar{\mathsf{u}}$ solves (4.4) if and only if

$$\bar{\mathsf{u}} = \max\left\{\mathsf{a}, \min\left\{\mathsf{b}, -\frac{1}{\lambda}\bar{\mathsf{p}}\right\}\right\}.$$

This immediately yields $\bar{\mathsf{u}} \in H^1(\varpi, \Omega)$ by invoking Kinderlehrer & Stampacchia (1980, Theorem A.1). □

Using this smoothness and an interpolation theorem between weighted spaces, we can bound the projection error in Corollary 4.1 and finish the error estimate (4.7) as follows.

THEOREM 4.3 (Rates of convergence) In the setting of Corollary 4.1, we have

$$\|\bar{\mathsf{u}} - \bar{\mathsf{u}}_h\|_{L^2(\Omega)} \lesssim \sigma_{\mathscr{T}} \left( \|\nabla \bar{\mathsf{p}}\|_{L^2(\varpi,\Omega)} + \|\nabla \bar{\mathsf{y}}\|_{L^\infty(\Omega)} \right), \tag{4.14}$$

where $\sigma_{\mathscr{T}}$ is defined in (4.6) and the hidden constant is independent of $\mathscr{T}$ and the continuous and discrete optimal pairs.

*Proof.* We need to bound the projection error $\|\bar{\mathsf{u}} - \Pi_{L^2}\bar{\mathsf{u}}\|_{L^2(\Omega)}$ only. Proposition 4.2 yields $\bar{\mathsf{u}} \in H^1(\varpi, \Omega)$; then, invoking Nochetto *et al.* (2016, Theorem 6.2), we derive

$$\|\bar{\mathsf{u}} - \Pi_{L^2}\bar{\mathsf{u}}\|_{L^2(\Omega)} \lesssim \sigma_{\mathscr{T}} \|\nabla \bar{\mathsf{u}}\|_{L^2(\varpi,\Omega)}.$$

Substituting the previous estimate in the conclusion of Corollary 4.1, we derive the claimed convergence rates. □

REMARK 4.4 (Rates of convergence for optimal control.) Estimate (4.14), for $n = 2$, is nearly optimal in terms of approximation. In contrast, in the three-dimensional case, the derived estimate (4.14) is suboptimal. However, the numerical experiment of Section 6.5 suggests that this is not sharp. The projection formula of Proposition 4.2 hints at the fact that the singularities of $\bar{\mathsf{p}}$ might not be present in $\bar{\mathsf{u}}$, which allows for a better rate of convergence.

On the basis of the previous results, we now derive an error estimate for the approximation of the state variable.

THEOREM 4.5 (Rates of convergence.) In the setting of Corollary 4.1 we have

$$\|\bar{\mathsf{y}} - \bar{\mathsf{y}}_h\|_{L^\infty(\Omega)} \lesssim \sigma_{\mathscr{T}} \left( \|\nabla \bar{\mathsf{p}}\|_{L^2(\varpi,\Omega)} + \|\nabla \bar{\mathsf{y}}\|_{L^\infty(\Omega)} \right), \tag{4.15}$$

where $\sigma_{\mathscr{T}}$ is defined in (4.6) and the hidden constant is independent of $\mathscr{T}$ and the continuous and discrete optimal pairs.

*Proof.* We start with a simple application of the triangle inequality:

$$\|\bar{\mathsf{y}} - \bar{\mathsf{y}}_h\|_{L^\infty(\Omega)} \leq \|\bar{\mathsf{y}} - \hat{\mathsf{y}}\|_{L^\infty(\Omega)} + \|\hat{\mathsf{y}} - \bar{\mathsf{y}}_h\|_{L^\infty(\Omega)}, \tag{4.16}$$

where $\hat{\mathsf{y}}$ solves $a(\hat{\mathsf{y}}, v) = (\bar{\mathsf{u}}_h, v)$ for all $v \in H_0^1(\Omega)$. The second term on the right-hand side of the previous inequality is controlled in view of standard pointwise estimates for finite elements. In fact, Schatz & Wahlbin (1982, Theorem 5.1) yields

$$\|\hat{\mathsf{y}} - \bar{\mathsf{y}}_h\|_{L^\infty(\Omega)} \lesssim h_{\mathscr{T}} |\log h_{\mathscr{T}}| \|\nabla \hat{\mathsf{y}}\|_{L^\infty(\Omega)}. \tag{4.17}$$

To control the first term on the right-hand side of (4.16), we invoke the same arguments that allowed us to conclude (4.12). We thus arrive at

$$\|\bar{\mathsf{y}} - \hat{\mathsf{y}}\|_{L^\infty(\Omega)} \lesssim \|\bar{\mathsf{u}} - \bar{\mathsf{u}}_h\|_{L^2(\Omega)}.$$

In view of (4.17), the previous estimate and the results of Theorem 4.3 allow us to derive the desired error estimates. □

REMARK 4.6 (Rates of convergence for the optimal state.) The error estimate (4.15), for $n = 2$, is near optimal in terms of regularity but suboptimal in terms of approximation. It relies on the $W^{1,\infty}(\Omega)$-regularity of the optimal state $\bar{\mathsf{y}}$ that solves problem (4.3); such a regularity property is guaranteed by references Maz′ya & Rossmann (1991), Fromm (1993) and Guzmán *et al.* (2009). The numerical experiments of Sections 6.2–6.4 suggest that, in the case that a better regularity for the optimal state is available, let us say $\bar{\mathsf{y}} \in W^{2,\infty}(\Omega)$, the order of convergence is quadratic.

## 5. Optimization with singular sources

Let us remark that, since the formulation of the adjoint problem (4.5) led to an elliptic problem with Dirac deltas on the right-hand side, the problem with point sources on the state (1.8)–(1.10) is, in a sense, dual to one with point observations (1.5)–(1.7). In the latter, the functional space for the adjoint variable is the one needed for the state variable in (1.8)–(1.10). The analysis will follow the one presented in Section 4.2. It is important to comment that problem (1.8)–(1.10) has been analysed before. We refer the reader to Gong *et al.* (2014b) for the elliptic case and to Seidman *et al.* (2012), Gong (2013), Leykekhman & Vexler (2013) and Gong *et al.* (2014a) for the parabolic one. It is our desire in this section to show how the theory of Muckenhoupt weights can be used to analyse and approximate problem (1.8)–(1.10). In doing this, it will be essential to assume that $\operatorname{dist}(\mathcal{D}, \partial\Omega) \geq d_{\mathcal{D}} > 0$. Set

- $\mathbb{H} = L^2(\Omega)$ and $C = \mathrm{id}$;
- $\mathbb{U} = \mathbb{R}^l$;
- $\mathbb{Y}_1 = H_0^1(\varpi, \Omega)$ and $\mathbb{X}_1 = H_0^1(\varpi^{-1}, \Omega)$, with $\varpi$ defined, as in Section 4.1, by (4.1);
- $\mathbb{Y}_2 = \mathbb{X}_2 = H_0^1(\Omega)$ and

$$a(v, w) = \int_\Omega \nabla v(x)^{\mathrm{T}} \nabla w(x)\, \mathrm{d}x.$$

- the bilinear form $b : \mathbb{U} \times (\mathbb{X}_1 + \mathbb{X}_2)$ to be

$$b(\mathbf{v}, w) = \sum_{z \in \mathcal{D}} v_z \langle \delta_z, w \rangle_{H_0^1(\varpi^{-1},\Omega)' \times H_0^1(\varpi^{-1},\Omega)}.$$

  Since, for $z \in \Omega$, $\delta_z \in H_0^1(\varpi^{-1}, \Omega)'$, we have that $b$ is continuous on $\mathbb{R}^l \times H_0^1(\varpi^{-1}, \Omega)$.

Let us now verify the assumptions of Theorem 2.2. The embedding of Nochetto *et al.* (2016, Lemma 7.6) yields that $y = Su \in \mathbb{Y}_1 = H_0^1(\varpi, \Omega) \hookrightarrow L^2(\Omega) = \operatorname{Dom}(C)$. The fact that $\mathbb{Y}_2 \subset \operatorname{Dom}(C)$ is trivial. Since $\Omega$ is convex, we invoke Nochetto *et al.* (2016, Lemma 7.6), again, and conclude that

$\bar{\mathsf{p}} \in H^2(\Omega) \cap H_0^1(\Omega) \hookrightarrow H_0^1(\varpi^{-1}, \Omega)$, which puts us in the setting of item (ii) with, once again, $\mathbb{D} = C_0^\infty(\Omega)$. Consequently, the optimality conditions hold.

For $\mathsf{a}, \mathsf{b} \in \mathbb{R}^l$ with $\mathsf{a}_z < \mathsf{b}_z$, we define the set of admissible controls as

$$\mathbb{U}_\delta = \left\{ \mathbf{u} \in \mathbb{R}^l : \mathsf{a}_z \leq u_z \leq \mathsf{b}_z \; \forall \, z \in \mathcal{D} \right\}.$$

The space of controls is already discrete, so we set $\mathbb{U}^h = \mathbb{U}$ and $\mathbb{U}^h_{\text{ad}} = \mathbb{U}_\delta$. Finally, we set, for $i = 1, 2$, $\mathbb{X}_i^h = \mathbb{Y}_i^h = \mathbb{V}(\mathscr{T})$, which, once again, trivializes (2.19) and the assumptions of Theorem 2.2. Since the bilinear form $b$ is not continuous on $\mathbb{U} \times \mathbb{X}_2$, we need to slightly modify the arguments of Lemma 2.3. In what follows, for $v \in C(\bar{\Omega})$ and $w \in \mathbb{R}^l$ we define

$$\langle v, w \rangle_{\mathcal{D}} := \sum_{z \in \mathcal{D}} v(z) w_z. \tag{5.1}$$

In this setting, the main error estimate for problem (1.8)–(1.10) is provided below. We comment that our proof is inspired by the arguments developed in Rannacher & Scott (1982), Leykekhman & Vexler (2013) and Gong *et al.* (2014b, Theorem 3.7).

THEOREM 5.1 (Rates of convergence) Let $\bar{\mathbf{u}}$ and $\bar{\mathbf{u}}_h$ be the continuous and discrete optimal controls, respectively, and assume that for every $q \in (2, \infty)$, $y_d \in L^q(\Omega)$. Let $\epsilon > 0$ and $\Omega_1$ be such that $\mathcal{D} \Subset \Omega_1 \Subset \Omega$. If $n = 2$, then

$$\|\bar{\mathbf{u}} - \bar{\mathbf{u}}_h\|_{\mathbb{R}^l} \lesssim h_{\mathscr{T}}^{2-\epsilon} \left( \|\bar{\mathbf{u}}\|_{\mathbb{R}^l} + \|\bar{\mathsf{p}}\|_{H^2(\Omega)} + \|\bar{\mathsf{p}}\|_{W^{2,r}(\Omega_1)} \right). \tag{5.2}$$

On the other hand, if $n = 3$, then

$$\|\bar{\mathbf{u}} - \bar{\mathbf{u}}_h\|_{\mathbb{R}^l} \lesssim h_{\mathscr{T}}^{1-\epsilon} \left( \|\bar{\mathbf{u}}\|_{\mathbb{R}^l} + \|\bar{\mathsf{p}}\|_{H^2(\Omega)} + \|\bar{\mathsf{p}}\|_{W^{2,r}(\Omega_1)} \right), \tag{5.3}$$

where $r < n/(n-2)$. The hidden constants in both estimates are independent of $\mathscr{T}$, and the continuous and discrete optimal pairs.

*Proof.* We start the proof by noticing that, since $\bar{\mathsf{y}} - y_d \in L^2(\Omega)$ and $\Omega$ is convex, standard regularity arguments (Grisvard, 1985) yield $\bar{\mathsf{p}} \in H^2(\Omega) \hookrightarrow C(\bar{\Omega})$ . This guarantees that pointwise evaluations of $\bar{\mathsf{p}}$ are well defined. Moreover, since, in this setting, $\mathbb{U}^h_{\text{ad}} = \mathbb{U}_{\text{ad}}$ estimate (2.21) reduces to

$$\lambda \|\bar{\mathbf{u}} - \bar{\mathbf{u}}_h\|^2_{\mathbb{R}^l} \leq \langle \bar{\mathsf{p}} - \bar{\mathsf{p}}_h, \bar{\mathbf{u}}_h - \bar{\mathbf{u}} \rangle_{\mathcal{D}},$$

where $\langle \cdot, \cdot \rangle_{\mathcal{D}}$ is defined in (5.1). Adding and subtracting the solution to (2.18) $\hat{\mathsf{p}}_h$, we obtain

$$\lambda \|\bar{\mathbf{u}} - \bar{\mathbf{u}}_h\|^2_{\mathbb{R}^l} \leq \langle \bar{\mathsf{p}} - \hat{\mathsf{p}}_h, \bar{\mathbf{u}}_h - \bar{\mathbf{u}} \rangle_{\mathcal{D}} + \langle \hat{\mathsf{p}}_h - \bar{\mathsf{p}}_h, \bar{\mathbf{u}}_h - \bar{\mathbf{u}} \rangle_{\mathcal{D}}. \tag{5.4}$$

This, in view of $\langle \hat{\mathsf{p}}_h - \bar{\mathsf{p}}_h, \bar{\mathbf{u}}_h - \bar{\mathbf{u}} \rangle_{\mathcal{D}} = -\|\hat{\mathsf{y}}_h - \bar{\mathsf{y}}_h\|^2_{L^2(\Omega)}$, implies that

$$\begin{aligned} \lambda \|\bar{\mathbf{u}} - \bar{\mathbf{u}}_h\|^2_{\mathbb{R}^l} + \|\hat{\mathsf{y}}_h - \bar{\mathsf{y}}_h\|^2_{L^2(\Omega)} &\leq \langle \bar{\mathsf{p}} - \hat{\mathsf{p}}_h, \bar{\mathbf{u}}_h - \bar{\mathbf{u}} \rangle_{\mathcal{D}} \\ &= \langle \bar{\mathsf{p}} - \mathsf{q}_h, \bar{\mathbf{u}}_h - \bar{\mathbf{u}} \rangle_{\mathcal{D}} + \langle \mathsf{q}_h - \hat{\mathsf{p}}_h, \bar{\mathbf{u}}_h - \bar{\mathbf{u}} \rangle_{\mathcal{D}}, \end{aligned} \tag{5.5}$$

where $\mathsf{q}_h$ is defined as the unique solution to

$$\mathsf{q}_h \in \mathbb{V}(\mathscr{T}): \quad a(w_h, \mathsf{q}_h) = (\bar{\mathsf{y}} - y_d, w_h)_{L^2(\Omega)} \quad \forall\, w_h \in \mathbb{V}(\mathscr{T}).$$

Since, by assumption, we have $d_{\mathcal{D}} > 0$, we can conclude that there are smooth subdomains $\Omega_0$ and $\Omega_1$ such that $\mathcal{D} \subset \Omega_0 \Subset \Omega_1 \Subset \Omega$. In view of (5.5), this key property will allow us to derive interior $L^\infty$ estimates for $\bar{\mathsf{p}} - \mathsf{q}_h$ and $\mathsf{q}_h - \hat{\mathsf{p}}_h$.

Let us first bound the difference $\bar{\mathsf{p}} - \mathsf{q}_h$. To do this, we notice that, since $\bar{\mathsf{y}} \in W_0^{1,s}(\Omega)$ for $s < n/(n-1)$, a standard Sobolev embedding result implies that $\bar{\mathsf{y}} \in L^r(\Omega)$ with $r \le ns/(n-s) < n/(n-2)$. Then, on the basis of the fact that $y_d \in L^q(\Omega)$ for $q < \infty$, interior regularity results guarantee that $\bar{\mathsf{p}} \in W^{2,r}(\Omega_1)$ for $r < n/(n-2)$. Consequently, since $\mathsf{q}_h$ corresponds to the Galerkin approximation of $\bar{\mathsf{p}}$, Schatz & Wahlbin (1977, Theorem 5.1) yields, when $n = 2$, that for any $\epsilon > 0$, we have

$$\|\bar{\mathsf{p}} - \mathsf{q}_h\|_{L^\infty(\Omega_0)} \lesssim \left(h_{\mathscr{T}}^{2-\epsilon} \|\bar{\mathsf{p}}\|_{W^{2,r}(\Omega_1)} + h_{\mathscr{T}}^2 \|\bar{\mathsf{p}}\|_{H^2(\Omega)}\right). \tag{5.6}$$

When $n = 3$, we have $\bar{\mathsf{p}} \in H_0^1(\Omega) \cap W^{2,r}(\Omega_1)$ for $r < 3$ and, as a consequence,

$$\|\bar{\mathsf{p}} - \mathsf{q}_h\|_{L^\infty(\Omega_0)} \lesssim \left(h_{\mathscr{T}}^{1-\epsilon} \|\bar{\mathsf{p}}\|_{W^{2,r}(\Omega_1)} + h_{\mathscr{T}}^2 \|\bar{\mathsf{p}}\|_{H^2(\Omega)}\right). \tag{5.7}$$

It remains then to estimate the difference $P_h = \mathsf{q}_h - \hat{\mathsf{p}}_h$. To do so, we employ a duality argument that combines the ideas of Rannacher & Scott (1982), Leykekhman & Vexler (2013) and Nochetto *et al.* (2016, Corollary 7.9). We start by defining $\varphi \in H_0^1(\Omega)$ as the solution to

$$a(v, \varphi) = \int_\Omega \operatorname{sgn}(\bar{\mathsf{y}} - \hat{\mathsf{y}}_h) v \quad \forall\, v \in H_0^1(\Omega), \tag{5.8}$$

where $\hat{\mathsf{y}}_h$ solves (2.17). Notice that $\|\operatorname{sgn}(\bar{\mathsf{y}} - \hat{\mathsf{y}}_h)\|_{L^\infty(\Omega)} \le 1$ for all $\mathscr{T} \in \mathbb{T}$. Therefore, Schatz & Wahlbin (1977, Theorem 5.1) followed by Schatz & Wahlbin (1982, Theorem 5.1) leads to (see also Gong *et al.*, 2014b, Lemma 3.2)

$$\|\varphi - \varphi_h\|_{L^\infty(\Omega_0)} \lesssim h_{\mathscr{T}}^2 |\log h_{\mathscr{T}}|^2, \tag{5.9}$$

where $\varphi_h$ is the Galerkin projection of $\varphi$ and the hidden constant does not depend on $\mathscr{T}$ or $\varphi$. In addition, we have $\varphi \in H^2(\Omega) \cap H_0^1(\Omega) \hookrightarrow H_0^1(\varpi^{-1}, \Omega)$ (Nochetto *et al.*, 2016, Lemma 7.6). Therefore $\varphi$ is a valid test function in the variational problem that $\bar{\mathsf{y}}$ solves. Then, using the continuity of the bilinear form $a$ and Galerkin orthogonality, we arrive at

$$\begin{aligned}
\|\bar{\mathsf{y}} - \hat{\mathsf{y}}_h\|_{L^1(\Omega)} &= \int_\Omega \operatorname{sgn}(\bar{\mathsf{y}} - \hat{\mathsf{y}}_h)(\bar{\mathsf{y}} - \hat{\mathsf{y}}_h) = a(\bar{\mathsf{y}} - \hat{\mathsf{y}}_h, \varphi) \\
&= a(\bar{\mathsf{y}}, \varphi - \varphi_h) = \langle \varphi - \varphi_h, \bar{\mathbf{u}} \rangle_{\mathcal{D}} \lesssim \|\bar{\mathbf{u}}\|_{\mathbb{R}^l} \|\varphi - \varphi_h\|_{L^\infty(\Omega_0)} \\
&\lesssim h_{\mathscr{T}}^2 |\log h_{\mathscr{T}}|^2 \|\bar{\mathbf{u}}\|_{\mathbb{R}^l},
\end{aligned}$$

where in the last step we used estimate (5.9).

We now recall that $P_h$ solves

$$a(w_h, P_h) = (\bar{\mathsf{y}} - \hat{\mathsf{y}}_h, w_h)_{L^2(\Omega)} \quad \forall\, w_h \in \mathbb{V}(\mathscr{T});$$

an inverse inequality and a stability estimate for the problem above yield

$$\|P_h\|^2_{L^\infty(\Omega)} \lesssim \mathfrak{i}^2_{\mathscr{T}} \|\nabla P_h\|^2_{L^2(\Omega)} \leq \mathfrak{i}^2_{\mathscr{T}} \|\bar{\mathsf{y}} - \hat{\mathsf{y}}_h\|_{L^1(\Omega)} \|P_h\|_{L^\infty(\Omega)},$$

where $\mathfrak{i}_{\mathscr{T}}$ is the mesh-dependent factor in the inverse inequality between $L^\infty(\Omega)$ and $H^1(\Omega)$ (see Brenner & Scott, 2008, Lemma 4.9.2 for $n = 2$ and Ern & Guermond, 2004, Lemma 1.142 for $n = 3$):

$$\mathfrak{i}_{\mathscr{T}} = (1 + |\log h_{\mathscr{T}}|)^{1/2} \quad \text{if} \quad n = 2, \quad \text{and} \quad \mathfrak{i}_{\mathscr{T}} = h_{\mathscr{T}}^{-1/2} \quad \text{if} \quad n = 3. \tag{5.10}$$

In conclusion,

$$\|P_h\|_{L^\infty(\Omega)} \lesssim \mathfrak{i}^2_{\mathscr{T}} \|\bar{\mathsf{y}} - \hat{\mathsf{y}}_h\|_{L^1(\Omega)} \lesssim \mathfrak{i}^2_{\mathscr{T}} h^2_{\mathscr{T}} |\log h_{\mathscr{T}}|^2 \|\bar{\mathbf{u}}\|_{\mathbb{R}^l}. \tag{5.11}$$

Combining the obtained pointwise bounds for $\bar{\mathsf{p}} - \mathsf{q}_h$ and $\mathsf{q}_h - \hat{\mathsf{p}}_h$, we obtain the desired estimates. □

REMARK 5.2 (Comparison with the literature) Reference Gong *et al.* (2014b) claims to obtain better rates than those in Theorem 5.1, namely, they can trade the term $h_{\mathscr{T}}^{-\epsilon}$ by a logarithmic factor $|\log h_{\mathscr{T}}|^s$ with $s \geq 1$ but small. However, when following the arguments that lead to this estimate (see Gong *et al.*, 2014b, formula (3.40)) one realizes that a slight inaccuracy takes place. Namely, the authors claim that, for $s < n/(n-1)$,

$$h_{\mathscr{T}}^{3-n/s} |\log h_{\mathscr{T}}| \lesssim h^2_{\mathscr{T}} |\log h_{\mathscr{T}}|.$$

However, $3 - n/s < 4 - n$ which, for $n = 2$ or $n = 3$, reduces to the estimates that we obtained in Theorem 5.1.

REMARK 5.3 (Rates of convergence for optimal control.) The error estimates (5.2) and (5.3) are suboptimal in terms of approximation; optimal error estimates should be be quadratic. In our method of proof, suboptimality is a consequence of estimates (5.6) and (5.7), which exploit the local regularity of the optimal adjoint state $\bar{\mathsf{p}}$ and estimate (5.11). Notice that the situation is worse for $n = 3$.

To conclude, we present an error estimate for the state variable.

COROLLARY 5.4 (Rates of convergence.) In the setting of Theorem 5.1 we have, for $n \in \{2, 3\}$,

$$\|\bar{\mathsf{y}} - \bar{\mathsf{y}}_h\|_{L^2(\Omega)} \lesssim \sigma_{\mathscr{T}} \left( \|\bar{\mathbf{u}}\|_{\mathbb{R}^l} + \|\bar{\mathsf{p}}\|_{H^2(\Omega)} + \|\bar{\mathsf{p}}\|_{W^{2,r}(\Omega_1)} + \|\nabla \bar{\mathsf{y}}\|_{L^2(\varpi,\Omega)} \right),$$

where the hidden constant is independent of $\mathscr{T}$ and the continuous and discrete optimal pairs.

*Proof.* A simple application of the triangle inequality yields

$$\|\bar{\mathsf{y}} - \bar{\mathsf{y}}_h\|_{L^2(\Omega)} \leq \|\bar{\mathsf{y}} - \hat{\mathsf{y}}_h\|_{L^2(\Omega)} + \|\hat{\mathsf{y}}_h - \bar{\mathsf{y}}_h\|_{L^2(\Omega)}, \tag{5.12}$$

where $\hat{\mathsf{y}}_h$ solves (2.17). To estimate the first term on the right-hand side of the previous expression, we invoke Nochetto *et al.* (2016, Corollary 7.9) and arrive at

$$\|\bar{\mathsf{y}} - \hat{\mathsf{y}}_h\|_{L^2(\Omega)} \lesssim \sigma_{\mathscr{T}} \|\nabla \bar{\mathsf{y}}\|_{L^2(\varpi,\Omega)}.$$

Using (5.5) and the results of Theorem 5.1, we bound the second term on the right-hand side of (5.12). This concludes the proof. □

## 6. Numerical experiments

In this section, we conduct a series of numerical experiments that illustrate the performance of the scheme (2.12)–(2.14) when it is used to approximate the solution to the optimization problem with point observations studied in Section 4 and the one with singular sources analysed in Section 5. Since, in general, it is rather difficult to find fundamental solutions, in some examples we modify the adjoint or state equations to versions where the solution is the restriction of the fundamental solution to the Poisson problem in the whole space to $\Omega$ and study the discretization of the ensuing system of equations. We are aware that this is not the optimality system of the problem, but it retains its essential difficulties and singularities and allows us to evaluate the rates of convergences.

### 6.1 *Implementation*

All the numerical experiments that will be presented have been carried out with the help of a code that is implemented using `C++`. The matrices involved in the computations have been assembled exactly, while the right-hand sides and the approximation errors are computed by a quadrature formula that is exact for polynomials of degree 19 for two-dimensional domains and degree 14 for three-dimensional domains. The corresponding linear systems are solved using the multifrontal massively parallel sparse direct solver (MUMPS) (Amestoy *et al.*, 2000, 2001). To solve the minimization problem (2.12)–(2.14) we use a Newton-type primal–dual active set strategy (Tröltzsch, 2010, Section 2.12.4).

For all our numerical examples, we consider $\lambda = 1$. We construct exact solutions based on the fundamental solutions for the Laplace operator:

$$\phi(x) = \begin{cases} -\dfrac{1}{2\pi} \displaystyle\sum_{z\in\mathfrak{S}} \log|x-z| & \text{if } \Omega = (0,1)^2 \subset \mathbb{R}^2, \\ \dfrac{1}{4\pi} \displaystyle\sum_{z\in\mathfrak{S}} \dfrac{1}{|x-z|} & \text{if } \Omega = (0,1)^3 \subset \mathbb{R}^3, \end{cases} \tag{6.1}$$

where, depending on the problem, $\mathfrak{S} = \mathcal{Z}$ or $\mathfrak{S} = \mathcal{D}$. We will also consider the fundamental solution of the Laplace operator in $\Omega = B_1 = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 < 1\} \subset \mathbb{R}^2$, which reads

$$\phi(x) = \frac{1}{2\pi} \log|x|. \tag{6.2}$$

We must remark that the introduction of weights is only to simplify the analysis and that these are never used in the implementation. This greatly simplifies it and allows for the use of existing codes.

To present the performance of the fully discrete scheme (2.12)–(2.14), we consider a family of quasi–uniform meshes $\{\mathscr{T}_k\}_{k=1}^8$. We set $N(k) = \#\mathscr{T}_k$, that is, the total number of degrees of freedom of the mesh $\mathscr{T}_k$. In addition, we denote by $\mathrm{EOC}_{\mathsf{q}}(k)$ the corresponding experimental order of convergence associated with the variable $\mathsf{q}$, which is computed using the formula

$$\mathrm{EOC}_{\mathsf{q}}(k) = \frac{\ln\left(e_{\mathsf{q}}(k-1)/e_{\mathsf{q}}(k)\right)}{\ln\left(N(k-1)/N(k)\right)},$$

where $e_{\mathsf{q}}(k)$ denotes the resulting error in the approximation of the variable $\mathsf{q}$ and $k \in \{2, \ldots, 8\}$.

TABLE 1 *Experimental order of convergence of scheme* (2.12)–(2.14) *when it is used to approximate the solution to the optimization problem of Section* 4 *with one observation point and* $\Omega = B_1$. *The* $\mathrm{EOC}_{\bar{\mathsf{u}}}$ *is in agreement with estimate* (4.14) *of Theorem* 4.3*: the family* $\{\mathscr{T}_k\}_{k=1}^{8}$ *is quasi-uniform and, thus,* $h_{\mathscr{T}_k} \approx N(k)^{-1/2}$*, which is what we observe. The* $\mathrm{EOC}_{\bar{\mathsf{y}}}$ *reveals quadratic order; see Remark* 4.6 *for a discussion*

| DOFs | $\\|\bar{\mathsf{u}} - \bar{\mathsf{u}}_{\mathscr{T}_k}\\|_{L^2(\Omega)}$ | $\mathrm{EOC}_{\bar{\mathsf{u}}}$ | $\\|\bar{\mathsf{y}} - \bar{\mathsf{y}}_{\mathscr{T}_k}\\|_{L^\infty(\Omega)}$ | $\mathrm{EOC}_{\bar{\mathsf{y}}}$ |
|---|---|---|---|---|
| 26 | 0.0595209 | – | 0.4816528 | – |
| 82 | 0.0359273 | −0.4395090 | 0.1656580 | −0.92919815 |
| 290 | 0.0175814 | −0.5657675 | 0.0442101 | −1.04576649 |
| 1090 | 0.0084497 | −0.5533850 | 0.0117083 | −1.00347662 |
| 4226 | 0.0043345 | −0.4926096 | 0.0030234 | −0.99914230 |
| 16642 | 0.0021736 | −0.5035636 | 0.0007708 | −0.99710702 |
| 66050 | 0.0010911 | −0.4999690 | 0.0002100 | −0.94329927 |
| 263170 | 0.0005472 | −0.4992283 | 0.0000567 | −1.02762135 |

### 6.2 *Optimization with point observations on a disk: one point*

We set $n = 2$ and $\Omega = B_1$. We set the control bounds that define the set $\mathbb{U}_{\mathcal{Z}}$ in (4.2) to $\mathsf{a} = -0.2$ and $\mathsf{b} = -0.1$. To construct an exact solution to the optimization problem with point observations, we slightly modify the corresponding state equation by adding a forcing term $\mathsf{f} \in L^2(\Omega)$, i.e., we replace (4.3) by the following problem:

$$\bar{\mathsf{y}} \in H_0^1(\Omega): \qquad a(\bar{\mathsf{y}}, w) = (\bar{\mathsf{u}} + \mathsf{f}, w) \quad \forall\, w \in H_0^1(\Omega). \tag{6.3}$$

We then define the exact optimal state, the observation set and the desired point value as follows:

$$\bar{\mathsf{y}}(x_1, x_2) = 2(1 - x^2 - y^2), \qquad \mathcal{Z} = \{(0,0)\}, \qquad \mathsf{y}_{(0,0)} = 1.$$

The exact optimal adjoint state is given by (6.2) and and the right-hand side $\mathsf{f}$ is computed accordingly. We notice that both $\bar{\mathsf{y}}$ and $\bar{\mathsf{p}}$ satisfy homogeneous Dirichlet boundary conditions.

Table 1 shows that, when approximating the optimal control variable, the $\mathrm{EOC}_{\bar{\mathsf{u}}}$ is in agreement with estimate (4.14). This illustrates the sharpness of the derived estimate up to a logarithmic term. We comment that, since the family $\{\mathscr{T}_k\}_{k=1}^{8}$ is quasi-uniform, we then have $h_{\mathscr{T}_k} \approx N(k)^{-1/2}$. Consequently, (4.14) reads as follows:

$$\\|\bar{\mathsf{u}} - \bar{\mathsf{u}}_{\mathscr{T}_k}\\|_{L^2(\Omega)} \lesssim N(k)^{-1/2} |\log N(k)|. \tag{6.4}$$

Table 1 also presents the $\mathrm{EOC}_{\bar{\mathsf{y}}}$ obtained for the approximation of the optimal state variable $\bar{\mathsf{y}}$: $h_{\mathscr{T}_k}^2 \approx N(k)^{-1}$; see Remark 4.6 for a discussion.

TABLE 2 *Experimental order of convergence of scheme* (2.12)–(2.14) *when it is used to approximate the solution to the optimization problem of Section 4 with one observation point and* $\Omega = (0,1)^2$. *The* $\mathrm{EOC}_{\bar{\mathsf{u}}}$ *is in agreement with estimate* (4.14) *of Theorem 4.3: the family* $\{\mathscr{T}_k\}$ *is quasi-uniform and, thus,* $h_{\mathscr{T}_k} \approx N(k)^{-1/2}$, *which is what we observe. The* $\mathrm{EOC}_{\bar{\mathsf{y}}}$ *reveals quadratic order; see Remark 4.6 for a discussion*

| DOFs | $\|\bar{\mathsf{u}} - \bar{\mathsf{u}}_{\mathscr{T}_k}\|_{L^2(\Omega)}$ | $\mathrm{EOC}_{\bar{\mathsf{u}}}$ | $\|\bar{\mathsf{y}} - \bar{\mathsf{y}}_{\mathscr{T}_k}\|_{L^\infty(\Omega)}$ | $\mathrm{EOC}_{\bar{\mathsf{y}}}$ |
|---|---|---|---|---|
| 42 | 0.0456202 | – | 0.3940558 | – |
| 146 | 0.0259039 | −0.4542396 | 0.1220998 | −0.9403796 |
| 546 | 0.0106388 | −0.6746618 | 0.0356279 | −0.9338121 |
| 2114 | 0.0053128 | −0.5129453 | 0.0104755 | −0.9042427 |
| 8322 | 0.0026798 | −0.4994327 | 0.0030256 | −0.9063059 |
| 33026 | 0.0013372 | −0.5043272 | 0.0008921 | −0.8860222 |
| 131586 | 0.0006675 | −0.5025385 | 0.0002586 | −0.8957802 |
| 525314 | 0.0003340 | −0.5000704 | 7.359881e-05 | −0.9077666 |

### 6.3 *Optimization with point observations on a square: one point*

We set $n = 2$, $\Omega = (0,1)^2$, $\mathsf{a} = -0.4$ and $\mathsf{b} = -0.2$. The state equation (4.3) is replaced by (6.3), which allows the incorporation of a forcing term $\mathsf{f}$. We then define the exact optimal state, the observation set and the desired point value as follows:

$$\bar{\mathsf{y}}(x_1, x_2) = 32x_1x_2(1-x_1)(1-x_2), \qquad \mathcal{Z} = \{(0.5, 0.5)\}, \qquad \mathsf{y}_{(0.5,0.5)} = 1.$$

The exact optimal adjoint state is given by (6.1) and the right-hand side $\mathsf{f}$ is computed accordingly. We notice that the optimal adjoint state $\bar{\mathsf{p}}$ does not satisfy homogeneous Dirichlet boundary conditions. We thus go beyond the theory developed in Section 4 and observe that, even if this is the case, Table 2 shows the optimal performance of the scheme (2.12)–(2.14) when approximating the solution to the optimization problem with point observations: $\mathrm{EOC}_{\bar{\mathsf{u}}}$ is in agreement with estimate (4.14) of Theorem 4.3.

### 6.4 *Optimization with point observations: four points*

The objective of this numerical experiment is to test the performance of the fully discrete scheme (2.12)–(2.14) when more observation points are considered in the optimization with point observations problem.

Let us consider $n = 2$ and $\Omega = (0,1)^2$. The control bounds defining the set $\mathbb{U}_{\mathcal{Z}}$ are given by $\mathsf{a} = -1.2$ and $\mathsf{b} = -0.7$. The state equation (4.3) is replaced by (6.3). This allows the incorporation of a forcing term $\mathsf{f}$. We set

$$\mathcal{Z} = \{(0.75, 0.75), (0.75, 0.25), (0.25, 0.75), (0.25, 0.25)\},$$

with corresponding desired values

$$\mathsf{y}_{(0.75,0.75)} = 1, \qquad \mathsf{y}_{(0.25,0.25)} = 1, \qquad \mathsf{y}_{(0.75,0.25)} = 0.5, \qquad \mathsf{y}_{(0.25,0.75)} = 0.5.$$

TABLE 3 *Experimental order of convergence of scheme* (2.12)–(2.14) *when it is used to approximate the solution of the problem of Section 4 with four observation points. The* $\mathrm{EOC}_{\bar{\mathsf{u}}}$ *is in agreement with estimate* (4.14) *of Theorem 4.3: the family* $\{\mathscr{T}_k\}$ *is quasi-uniform, so that* $h_{\mathscr{T}_k} \approx N(k)^{-1/2}$, *which is what we observe. The* $\mathrm{EOC}_{\bar{\mathsf{y}}}$ *reveals quadratic order; see Remark 4.6 for a discussion*

| DOFs | $\|\bar{\mathsf{u}} - \bar{\mathsf{u}}_{\mathscr{T}_k}\|_{L^2(\Omega)}$ | EOC | $\|\bar{\mathsf{y}} - \bar{\mathsf{y}}_{\mathscr{T}_k}\|_{L^\infty(\Omega)}$ | EOC |
|---|---|---|---|---|
| 42 | 0.0285416 | – | 0.0595256 | – |
| 146 | 0.0285084 | −0.0009357 | 0.0152388 | −1.0936039 |
| 546 | 0.0208153 | −0.2384441 | 0.0039226 | −1.0288683 |
| 2114 | 0.0116163 | −0.4308717 | 0.0010313 | −0.9868631 |
| 8322 | 0.0061821 | −0.4602926 | 0.0002708 | −0.9758262 |
| 33026 | 0.0030792 | −0.5056447 | 7.057710e-05 | −0.9755383 |
| 131586 | 0.0014908 | −0.5247299 | 1.729492e-05 | −1.0173090 |
| 525314 | 0.0007618 | −0.4849766 | 4.503108e-06 | −0.9720511 |

The exact optimal state variable is then given by

$$\bar{\mathsf{y}}(x_1, x_2) = 2.75 - 2x_1 - 2x_2 + 4x_1x_2$$

and the exact optimal adjoint state is given by (6.1). We must immediately comment that, as in the example of Section 6.3, both $\bar{\mathsf{y}}$ and $\bar{\mathsf{p}}$ do not satisfy homogeneous Dirichlet boundary conditions. However, as the results of Table 3 show, the $\mathrm{EOC}_{\bar{\mathsf{u}}}$ is optimal and in agreement with estimate (4.14) of Theorem 4.3. This illustrates the robustness of scheme (2.12)–(2.14) when more observations points are considered. Table 3 also shows quadratic order for the $\mathrm{EOC}_{\bar{\mathsf{y}}}$; see Remark 4.6 for a discussion.

### 6.5 *Optimization with point observations: a three-dimensional example*

We set $n = 3$ and $\Omega = (0, 1)^3$. We define the control bounds for the set $\mathbb{U}_{\mathcal{Z}}$ as follows: $\mathsf{a} = -15$ and $\mathsf{b} = -5$. The optimal state is

$$\bar{\mathsf{y}}(x_1, x_2, x_3) = \frac{8192}{27} x_1x_2x_3(1 - x_1)(1 - x_2)(1 - x_3),$$

whereas the optimal adjoint state is defined by (6.1). The set of observation points is

$$\mathcal{Z} = \{(0.25, 0.25, 0.25), (0.25, 0.25, 0.75), (0.25, 0.75, 0.25), (0.25, 0.75, 0.75), \\ (0.75, 0.25, 0.25), (0.75, 0.25, 0.75), (0.75, 0.75, 0.25), (0.75, 0.75, 0.25)\}$$

and we set $\mathsf{y}_z = 1$ for all $z \in \mathcal{Z}$. In this example, the optimal adjoint state $\bar{\mathsf{p}}$ does not satisfy homogeneous Dirichlet boundary conditions. However, as shown in Table 4, the performance of the scheme (2.12)–(2.16) is better than expected: $\mathrm{EOC}_{\bar{\mathsf{u}}}$ presents a better result than estimate (4.14) of Theorem 4.3; see Remark 4.4 for a discussion. We notice that since the family $\{\mathscr{T}_k\}_{k=1}^{8}$ is quasi-uniform, we have $h_{\mathscr{T}_k} \approx N(k)^{-1/3}$.

TABLE 4 *Experimental order of convergence of scheme* (2.12)–(2.14) *when it is used to approximate the solution to the optimization problem of Section 4 in a three-dimensional example. The* $\mathrm{EOC}_{\bar{\mathsf{u}}}$ *suggests that estimate* (4.14) *of Theorem 4.3 is not sharp; see Remark 4.4 for a discussion. We notice that the family* $\{\mathscr{T}_k\}$ *is quasi-uniform and then* $h_{\mathscr{T}_k} \approx N(k)^{-1/3}$

| DOFs | $\|\bar{\mathsf{u}} - \bar{\mathsf{u}}_{\mathscr{T}_k}\|_{L^2(\Omega)}$ | $\mathrm{EOC}_{\bar{\mathsf{u}}}$ |
|---|---|---|
| 1419 | 0.0274726 | – |
| 3694 | 0.0199406 | −0.3349167 |
| 9976 | 0.0120137 | −0.5100352 |
| 27800 | 0.0088690 | −0.2961201 |
| 79645 | 0.0067903 | −0.2537367 |
| 234683 | 0.0049961 | −0.2839348 |
| 704774 | 0.0037908 | −0.2510530 |
| 2155291 | 0.0028947 | −0.2412731 |

TABLE 5 *Experimental order of convergence of scheme* (2.12)–(2.14) *when it is used to approximate the solution to the optimization problem with point sources of Section 5. The* $\mathrm{EOC}_{\bar{\mathsf{u}}}$ *reveals quadratic order and illustrates our error estimate* (5.2)

| DOFs | $\|\bar{\mathsf{u}} - \bar{\mathsf{u}}_{\mathscr{T}_k}\|_{L^2(\Omega)}$ | EOC |
|---|---|---|
| 86 | 0.0536485 | – |
| 294 | 0.0207101 | −0.7743303 |
| 1094 | 0.0068950 | −0.8369949 |
| 4230 | 0.0021408 | −0.8648701 |
| 16646 | 0.0006380 | −0.8836678 |
| 66054 | 0.0001850 | −0.8981934 |
| 263174 | 5.259841e-05 | −0.9098104 |
| 1050630 | 1.472536e-05 | −0.9196613 |

### 6.6 *Optimization with singular sources*

We now explore the performance of scheme (2.12)–(2.14) when it is used to solve the optimization problem with singular sources. We set $n = 2$ and $\Omega = (0,1)^2$. We consider $\mathcal{D} = (0.5, 0.5)$ and the control bounds that define the set $\mathbb{U}_\delta$ are $\mathsf{a} = 0.3$ and $\mathsf{b} = 0.7$. The desired state and the exact adjoint state are defined as

$$\bar{\mathsf{p}}(x_1, x_2) = -32x_1x_2(1-x_1)(1-x_2), \qquad \bar{\mathsf{y}}_d = -\sin(2\pi x)\cos(2\pi x).$$

The exact optimal state is given by (6.1). We notice that the optimal state $\bar{\mathsf{y}}$ does not satisfy homogeneous Dirichlet boundary conditions; nevertheless, we explore the performance of (2.12)–(2.14) beyond the scope of the theory. As Table 5 shows, the experimental order of convergence $\mathrm{EOC}_{\bar{\mathsf{u}}}$ is optimal in terms of approximation.

**Acknowledgements**

The authors would like to thank Johnny Guzmán for fruitful discussions regarding pointwise estimates and the regularity of elliptic problems in convex, polytopal domains. We would also like to thank Alejandro Allendes for his technical support. Thanks also to the referees for their insightful comments and suggestions.

**Funding**

NSF (DMS-1521590 to H.A.); CONICYT through FONDECYT (3160201 to E.O.); NSF (DMS-1418784 to A.J.S.).

REFERENCES

ACHDOU, Y. (2005) An inverse problem for a parabolic variational inequality arising in volatility calibration with American options. *SIAM J. Control Optim.*, **43**, 1583–1615.

AGNELLI, J., GARAU, E. & MORIN, P. (2014) *A posteriori* error estimates for elliptic problems with Dirac measure terms in weighted spaces. *ESAIM Math. Model. Numer. Anal.*, **48**, 1557–1581.

AIMAR, H., CARENA, M., DURÁN, R. & TOSCHI, M. (2014) Powers of distances to lower dimensional sets as Muckenhoupt weights. *Acta Math. Hungar.*, **143**, 119–137.

AMESTOY, P., DUFF, I. & L'EXCELLENT, J.-Y. (2000) Multifrontal parallel distributed symmetric and unsymmetric solvers. *Comput. Methods Appl. Mech. Eng.*, **184**, 501–520.

AMESTOY, P., DUFF, I., L'EXCELLENT, J.-Y. & KOSTER, J. (2001) A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM J. Matrix Anal. Appl.*, **23**, 15–41 (electronic).

ANTIL, H. & OTÁROLA, E. (2015) A FEM for an optimal control problem of fractional powers of elliptic operators. *SIAM J. Control Optim.*, **53**, 3432–3456.

ANTIL, H., OTÁROLA, E. & SALGADO, A. J. (2016) A space-time fractional optimal control problem: analysis and discretization. *SIAM J. Control Optim.*, **54**, 1295–1328.

BERMÚDEZ, A., GAMALLO, P. & RODRÍGUEZ, R. (2004) Finite element methods in local active control of sound. *SIAM J. Control Optim.*, **43**, 437–465.

BRENNER, S. & SCOTT, L. (2008) *The Mathematical Theory of Finite Element Methods*, 3rd edn. Texts in Applied Mathematics, vol. 15. New York: Springer, pp. xviii+397.

BRETT, C., DEDNER, A. & ELLIOTT, C. (2016) Optimal control of elliptic PDEs at points. *IMA J. Numer. Anal.*, **36**, 1015–1050.

BRETT, C., ELLIOTT, C., HINTERMÜLLER, M. & LÖBHARD, C. (2015) Mesh adaptivity in optimal control of elliptic variational inequalities with point-tracking of the state. *Interfaces Free Bound.*, **17**, 21–53.

BURNS, J. & ITO, K. (1995) On well-posedness of integro-differential equations in weighted $L^2$-spaces. *Differential Integral Equations*, **8**, 627–646.

CAFFARELLI, L. & SILVESTRE, L. (2007) An extension problem related to the fractional Laplacian. *Comm. Part. Diff. Eqs.*, **32**, 1245–1260.

CANNARSA, P., MARTINEZ, P. & VANCOSTENOBLE, J. (2008) Carleman estimates for a class of degenerate parabolic operators. *SIAM J. Control Optim.*, **47**, 1–19.

CASAS, E. (1985) $L^2$ estimates for the finite element method for the Dirichlet problem with singular data. *Numer. Math.*, **47**, 627–632.

CASAS, E., CLASON, C. & KUNISCH, K. (2012) Approximation of elliptic control problems in measure spaces with sparse solutions. *SIAM J. Control Optim.*, **50**, 1735–1752.

CAVALHEIRO, A. (2011) A theorem on global regularity for solutions of degenerate elliptic equations. *Commun. Math. Anal.*, **11**, 112–123.

CHANG, L., GONG, W. & YAN, N. (2015) Numerical analysis for the approximation of optimal control problems with pointwise observations. *Math. Methods Appl. Sci.*, **38**, 4502–4520.

DAUGE, M. (1992) Neumann and mixed problems on curvilinear polyhedra. *Integral Equations Operator Theory*, **15**, 227–261.

DE LOS REYES, J. C. (2015) *Numerical PDE-Constrained Optimization*. SpringerBriefs in Optimization, Cham: Springer, pp. x+123.

DU, R. (2014) Approximate controllability of a class of semilinear degenerate systems with boundary control. *J. Differential Equations*, **256**, 3141–3165.

DUOANDIKOETXEA, J. (2001) *Fourier Analysis*. Graduate Studies in Mathematics, vol. 29. Providence, RI: American Mathematical Society, pp. xviii+222.

ERN, A. & GUERMOND, J.-L. (2004) *Theory and Practice of Finite Elements*. Applied Mathematical Sciences, vol. 159. New York: Springer, pp. xiv+524.

FABES, E., KENIG, C. & SERAPIONI, R. (1982) The local regularity of solutions of degenerate elliptic equations. *Comm. Part. Diff. Eqs.*, **7**, 77–116.

FABIANO, R. (2013) A semidiscrete approximation scheme for neutral delay-differential equations. *Int. J. Numer. Anal. Model.*, **10**, 712–726.

FABIANO, R. & TURI, J. (2003) Making the numerical abscissa negative for a class of neutral equations. *Discrete Contin. Dyn. Syst.*, 256–262. Dynamical systems and differential equations (Wilmington, NC, 2002).

FORNASIER, M., PICCOLI, B. & ROSSI, F. (2014) Mean-field sparse optimal control. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, **372**, 20130400, 21.

FROMM, S. (1993) Potential space estimates for Green potentials in convex domains. *Proc. Amer. Math. Soc.*, **119**, 225–233.

FULLER, C., NELSON, P. & ELLIOTT, S. (1996) *Active Control of Vibration*. London, San Diego: Academic Press.

GAMALLO, P. & HERNÁNDEZ, E. (2009) Error estimates for the approximation of a class of optimal control systems governed by linear PDEs. *Numer. Funct. Anal. Optim.*, **30**, 523–547.

GONG, W. (2013) Error estimates for finite element approximations of parabolic equations with measure data. *Math. Comp.*, **82**, 69–98.

GONG, W., HINZE, M. & ZHOU, Z. (2014a) A priori error analysis for finite element approximation of parabolic optimal control problems with pointwise control. *SIAM J. Control Optim.*, **52**, 97–119.

GONG, W., WANG, G. & YAN, N. (2014b) Approximations of elliptic optimal control problems with controls acting on a lower dimensional manifold. *SIAM J. Control Optim.*, **52**, 2008–2035.

GRISVARD, P. (1985) *Elliptic Problems in Nonsmooth Domains*. Monographs and Studies in Mathematics, vol. 24. Boston, MA: Pitman (Advanced Publishing Program), pp. xiv+410.

GUEYE, M. (2014) Exact boundary controllability of 1-D parabolic and hyperbolic degenerate equations. *SIAM J. Control Optim.*, **52**, 2037–2054.

GUZMÁN, J., LEYKEKHMAN, D., ROSSMANN, J. & SCHATZ, A. (2009) Hölder estimates for Green's functions on convex polyhedral domains and their applications to finite element methods. *Numer. Math.*, **112**, 221–243.

HERNÁNDEZ, E., KALISE, D. & OTÁROLA, E. (2010) Numerical approximation of the LQR problem in a strongly damped wave equation. *Comput. Optim. Appl.*, **47**, 161–178.

HERNÁNDEZ, E. & OTÁROLA, E. (2009) A locking-free FEM in active vibration control of a Timoshenko beam. *SIAM J. Numer. Anal.*, **47**, 2432–2454.

HINTERMÜLLER, M. & LAURAIN, A. (2008) Electrical impedance tomography: from topology to shape. *Control Cybernet.*, **37**, 913–933.

HINZE, M. (2005) A variational discretization concept in control constrained optimization: the linear-quadratic case. *Comput. Optim. Appl.*, **30**, 45–61.

HINZE, M., PINNAU, R., ULBRICH, M. & ULBRICH, S. (2009) *Optimization with PDE Constraints*. Mathematical Modelling: Theory and Applications, vol. 23. New York: Springer, pp. xii+270.

ITO, K. & KUNISCH, K. (2008) *Lagrange Multiplier Approach to Variational Problems and Applications*. Advances in Design and Control, vol. 15. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), pp. xviii+341.

ITO, K. & TURI, J. (1991) Numerical methods for a class of singular integro-differential equations based on semigroup approximation. *SIAM J. Numer. Anal.*, **28**, 1698–1722.

JERISON, D. & KENIG, C. (1981) The Neumann problem on Lipschitz domains. *Bull. Amer. Math. Soc. (N.S.)*, **4**, 203–207.

JERISON, D. & KENIG, C. (1995) The inhomogeneous Dirichlet problem in Lipschitz domains. *J. Funct. Anal.*, **130**, 161–219.

KINDERLEHRER, D. & STAMPACCHIA, G. (1980) *An Introduction to Variational Inequalities and Their Applications*. Pure and Applied Mathematics, vol. 88. New York-London: Academic Press. [Harcourt Brace Jovanovich, Publishers], pp. xiv+313.

LEYKEKHMAN, D., MEIDNER, D. & VEXLER, B. (2013) Optimal error estimates for finite element discretization of elliptic optimal control problems with finitely many pointwise state constraints. *Comput. Optim. Appl.*, **55**, 769–802.

LEYKEKHMAN, D. & VEXLER, B. (2013) Optimal a priori error estimates of parabolic optimal control problems with pointwise control. *SIAM J. Numer. Anal.*, **51**, 2797–2821.

LIONS, J.-L. (1971) *Optimal Control of Systems Governed by Partial Differential Equations.* Die Grundlehren der mathematischen Wissenschaften, Band 170. New York-Berlin: Springer, pp. xi+396.

MAZ′YA, V. & ROSSMANN, J. (1991) On the Agmon-Miranda maximum principle for solutions of elliptic equations in polyhedral and polygonal domains. *Ann. Global Anal. Geom.*, **9**, 253–303.

MAZ'YA, V. & ROSSMANN, J. (2010) *Elliptic Equations in Polyhedral Domains*. Mathematical Surveys and Monographs, vol. 162. Providence, RI: American Mathematical Society, pp. viii+608.

NELSON, P. & ELLIOTT, S. (1992) *Active Control of Sound*. London, San Diego: Academic Press.

NOCHETTO, R., OTÁROLA, E. & SALGADO, A. (2015) A PDE approach to fractional diffusion in general domains: a priori error analysis. *Found. Comput. Math.*, **15**, 733–791.

NOCHETTO, R., OTÁROLA, E. & SALGADO, A. (2016) Piecewise polynomial interpolation in Muckenhoupt weighted Sobolev spaces and applications. *Numer. Math.*, **132**, 85–130.

PIEPER, K. & VEXLER, B. (2013) A priori error analysis for discretization of sparse elliptic optimal control problems in measure space. *SIAM J. Control Optim.*, **51**, 2788–2808.

RANNACHER, R. & SCOTT, R. (1982) Some optimal error estimates for piecewise linear finite element approximations. *Math. Comp.*, **38**, 437–445.

RANNACHER, R. & VEXLER, B. (2005) A priori error estimates for the finite element discretization of elliptic parameter identification problems with pointwise measurements. *SIAM J. Control Optim.*, **44**, 1844–1863.

SAVARÉ, G. (1998) Regularity results for elliptic equations in Lipschitz domains. *J. Funct. Anal.*, **152**, 176–201.

SCHATZ, A. H. & WAHLBIN, L. B. (1977) Interior maximum norm estimates for finite element methods. *Math. Comp.*, **31**, 414–442.

SCHATZ, A. & WAHLBIN, L. (1982) On the quasi-optimality in $L_\infty$ of the $\dot{H}^1$-projection into finite element spaces. *Math. Comp.*, **38**, 1–22.

SCOTT, R. (1973/74) Finite element convergence for singular data. *Numer. Math.*, **21**, 317–327.

SEIDMAN, T., GOBBERT, M., TROTT, D. & KRUŽÍK, M. (2012) Finite element approximation for time-dependent diffusion with measure-valued source. *Numer. Math.*, **122**, 709–723.

TRÖLTZSCH, F. (2010) *Optimal Control of Partial Differential Equations*. Graduate Studies in Mathematics, vol. 112. Providence, RI: American Mathematical Society, pp. xvi+399.

TURESSON, B. O. (2000) *Nonlinear Potential Theory and Weighted Sobolev Spaces*. Lecture Notes in Mathematics, vol. 1736. Berlin: Springer, pp. xiv+173.

UNGER, A. & TRÖLTZSCH, F. (2001) Fast solution of optimal control problems in the selective cooling of steel. *ZAMM Z. Angew. Math. Mech.*, **81**, 447–456.